<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bentivogli, C.</td>
<td>Taxi regulation and the Bersani reform: a survey of major Italian cities</td>
</tr>
<tr>
<td>Shrivastava, P. O’Mahony, M.</td>
<td>Modeling an Integrated Public Transportation System - a case study in Dublin, Ireland</td>
</tr>
<tr>
<td>van Geenhuizen, M.</td>
<td>ICT applications on the road to sustainable urban transport</td>
</tr>
<tr>
<td>Kalantari, J. Sternberg, H.</td>
<td>Research Outlook on a Mixed Model Transportation Network</td>
</tr>
<tr>
<td>Timilsina, G.R. Dulal, H.B.</td>
<td>Regulatory instruments to control environmental externalities from the transport sector</td>
</tr>
</tbody>
</table>
Taxi regulation and the Bersani reform:
a survey of major Italian cities

Chiara Bentivogli

Banca d’Italia

Abstract

Using data from a Bank of Italy survey, this paper analyzes the Italian taxi market and its recent changes. Local regulations are rather homogeneous, while there is a widespread disproportion, within municipalities advisory committees, between the number of taxi drivers representatives and that of consumers’; indicators of service adequacy are seldom used. Service costs are rather homogeneous across Italian provinces, while there is great variance as to supply and fares. The instruments provided to municipalities by the new Bersani law have been used mainly in major cities. Service increase, achieved mainly through additional shifts rather than through the provision of (free) additional licenses, was often obtained in exchange for fare increases; the use of traffic policies has been almost absent. It is difficult to evaluate the adequacy of local decisions, given the lack of non-occasional information on market structure.

Keywords: Taxi; Regulation; Transport.

1. Introduction

As in many other branches of the transport sector, taxi and limousine services (in Italian “noleggio con conducente” – NCC) received recently the attention of the Italian legislator. Even if there are still a sizable regulation and high entry barriers in taxi markets, in the summer of 2006 a new legislation has given to municipalities additional instruments to increase supply.
This paper examines the situation, using data from a Bank of Italy survey in the first months of 2007 in 66 main Italian cities. The survey follows the first one carried out in 2004 in the capitals of Italian provinces (Bentivogli, Calderini, 2007), and investigates the links between the choices of the local regulator and the supply and efficiency of the service.

Section 2 summarizes the main characteristics of the service and the different kinds of regulation suggested by economic theory; section 3 presents the national regulatory framework. Sections 4 and 5 analyze the results of the second Bank of Italy’s survey on the differences among local regulations and taxi and NCC supply at the territorial level; finally section 6 presents some policy implications.

2. Service characteristics and regulation

From a technological point of view, taxi and NCC services show low fixed costs, sunk costs nearly null, low economies of scale (with the exception of the collateral services of radio-dispatching) and low economies of scope. The cost function, that includes the cost of labour (with very low professional requirements) and the costs of buying and operating the car (the cost of fuel in particular), is greatly influenced by a set of context variables (congestion, reserved lanes, urban density). These same factors can also influence demand, that depends – besides general economic context – on service cost and quality compared to other transport options.

Local transport demand - including taxi, limousine services as well as public and private transport - meets the needs of mobility of residents and of visitors of a city, being a derived demand. In general, the spatial movement of the consumer does not have an utility in itself, but allows her to benefit from goods and services otherwise not available. Consumer’s decisions include the destination, the timing, the route, and the transportation mode; the choice of a specific means of transport depends on relative monetary and time costs as well as on the relative intrinsic comfort. For a given price of transportation, the substitutability among the different transport modes depends on several factors: waiting time, relative speed and its uncertainty (variability), private cars access to the different city zones, availability and cost of parking, travel comfort, special needs (disabled persons, heavy luggage, and so on). In general taxi offers a point to point service more comfortable and faster than other means of transportation and without accessory monetary and time costs (for example, it does not require to find and pay for parking), usually with a lower uncertainty on expected transport time. High taxi waiting time and a surface traffic particularly severe can nevertheless make the underground (in the cities where it is available) a strict substitute for taxi.

Traffic control measures – for example the presence in large parts of the city of reserved lanes and parking only for taxi or public transit – can therefore reduce taxi costs and at the same time increase taxi demand. This shows how market equilibria can vary greatly – since they do not move along traditional demand and supply curves that relate the service price and quantity – depending on the contextual factors. More generally the nature of taxi demand, that depends on its price but also on a variety of service characteristics, creates the possibility of multiple equilibria: demand may in fact depend on supply size through its dependence on waiting time (which in its turn depends on supply as well as on the size of demand itself), so that an increase in supply
(for a given fare), by reducing waiting time, can increase demand (the so called Mohring effect).

This point can be clarified through an extremely simplified scheme (Douglas, 1972; Beesley, Glaister, 1983; Cairns, Liston-Heyes, 1996; Lam, Bell, 2003) where in each time unit (for example one hour) there are \( n \) taxi circulating in the city (\( n=Nh/24 \), where \( N \) is total licences/taxis number and \( h \) is the average number of working hours of taxi drivers) of which \( E \) are engaged and \( V \) are vacant (\( n=E+V \)). The demand for trips in each hour\(^2\) depends on the average fare per trip \( p \), on the waiting time of customers \( w \), on a series of exogenous variables \( x \), among which, for example, the average taxi speed (which is a function of the length of reserved lanes, and so on), and on the level of these variables with respect to that of alternative means of transport (same variables with an \( a \) index):

\[
Q = f\left(\frac{p}{p_a}, \frac{w}{w_a}, x\right) \quad p' < 0, w' < 0, x' > 0
\]  

where \( w \) is waiting time (expressed as a fraction of one hour) which depends on the number of vacant cabs \( V \), and \( p \) (the average fare per trip) is a function of current fares and of the average trip time \( t \) (fraction of an hour). In every point of time the average number of engaged and vacant cabs will be given by:

\[
E = tQ \quad (2)
\]
\[
V = n - tQ \quad (3)
\]
\[
w = w(V) \quad w' < 0 \quad (4)
\]

For given values of the remaining parameters, it is possible to write the waiting time of the customer as a decreasing function of supply (number of licences and average time of use of cabs) and as an increasing function of demand:

\[
w = w\left(Q, t, \frac{Nh}{24}\right) \quad (4')
\]

Under the assumption that waiting time elasticity with respect to vacant cabs is constant and equal to one for any level of \( V \), for example if \( w=g/V \), it can be rewritten as:

\[
w = \frac{g}{n - tf(p, w)} \quad g > 0 \quad (4'')
\]

This formulation helps to clarify that changes in supply (through changes in \( N \) and/or in \( h \), that represent respectively the extensive and intensive margin of supply change)

\(^2\) We assume that demand is uniformly distributed during the day.
and in fares – in Italy both fixed by the regulator – do not need necessarily to move in opposite directions. Even in the absence of exogenous shifts in demand curve – due to the local economic context, to the traffic management, and to the quality of taxi compared to that of other transportation services – an increase in supply (an increase in \( n \)), will reduce \( w \) and increase demand, without implying necessarily a reduction of \( p \).

The size of demand change will depend on the relationship between \( V \), \( w \) and \( E \):

\[
\frac{\partial E}{\partial n} = \frac{1}{1 - (V / (E \omega))} > 0 \quad \text{where} \quad \omega = \frac{w \frac{\partial Q}{Q} \frac{\partial Q}{Q}}{w} < 0 \tag{5}
\]

is the waiting time elasticity of demand for trips. If demand elasticity is low with respect to waiting time (\(|\omega| < 1\)), an increase in the number of cabs (\( n \)) will generate a less than proportional increase in demand. A low elasticity of demand with respect to waiting time could reflect for example a lack of substitutes among other means of local public transportation, due to a significant fare gap and a low availability of fast public transportation (underground or surface transit with reserved lanes). On the other hand, if demand is very elastic with respect to waiting time, then (for given \( p \)) an increase in the number of cabs \( n \) will create by itself a sufficient increase in demand.

Assuming that the hourly cost of taxi service is constant and independent of the cab being vacant or engaged,\(^3\) the impact of an increase in \( n \) on unit profits will depend only on demand sensitivity:

\[
\frac{\partial \pi}{\partial n} = \frac{\partial}{\partial n} \left( \frac{Ep}{tn^2} \right) = \frac{p}{tn^2} \left( \frac{\partial E}{\partial n} \frac{n}{E} - 1 \right) < 0 \quad \text{if} \quad \frac{\partial E}{\partial n} \frac{n}{E} < 1 \tag{6}
\]

If the function \( w \) is \((4'')\), (6) can be rewritten as:

\[
\frac{\partial \pi}{\partial n} = \frac{p}{tn^2} \left( \frac{1}{1 - V / (E \omega)} \frac{n}{E} - 1 \right) < 0 \quad \text{if} \quad |\omega| < 1 \tag{6'}
\]

The possibility of a relatively inelastic demand with respect to waiting time could explain the strong opposition of taxi drivers to an increase in the number of licences for given fares (see further on). Whilst the interaction between demand and service availability does not imply necessarily a decrease in unit profits (and even less lower fares that, in the strongly regulated settings that characterize the sector, are predetermined too), the incumbents’ perception is that demand in not very much sensitive to waiting time and hence to service availability.

On the other hand, the pressure of incumbents in favour of a fare increase implies that they perceive a low sensitivity of demand to fare changes. The impact on unit revenues of a fare increase, given \( n \), will be given by:

\[^3\] In short this assumption implies that the main component of taxi costs is the opportunity-cost of labour. Notice that in this case supply changes due to an increase in the number of licences \( N \) or to an increase of the intensity of taxi use (\( h/24\), where \( h \) is the average daily schedule of use of each taxi) will be quite equivalent.
\[
\frac{\partial \pi}{\partial p} = \frac{E}{tn} \left( 1 + \frac{\eta}{1 - \omega(E/V)} \right) > 0 \quad \text{if } 1 + \eta > \omega E/V
\]  
(7)

where \( \eta = \frac{\partial Q}{\partial p} \frac{p}{tE} < 0 \) is the price elasticity of demand for trips. Since the right hand side of the inequality (7) is always negative (\( \omega < 0 \)), unit revenues will increase for \( |\eta| \leq 1 \), while they could increase or fall if demand is very elastic with respect to price. In short, a rise in \( p \) produces a fall in demand but also in waiting time and this last variable tends in its turn to influence demand positively. More in general, even if the regulator chooses the level both of \( p \) and of \( n \), market equilibrium is determined by the value assumed by waiting time according to demand characteristics and its relationship with supply and demand (the 4th).

The role of waiting time as a variable that adjust the market to equilibrium explains why prices and quantities does not move always and necessarily in opposite directions (as it would be natural to expect given a normal demand schedule). In a competitive market - with free entry in the presence of extra profits and variations in service quantity depending on profitability conditions of the individual operator - price and waiting time will adjust so as to ensure a service availability suitable to demand. In a regulated market, waiting time and the actual use of the service (paying demand) will adjust given fares and service availability (in terms of \( N \) and \( h \), usually both fixed by the regulator).

Actually this interaction between demand and supply could also motivate the presence of a regulation in an industry that does not seem to have any feature of a natural monopoly: in a perfectly competitive market the suppliers of the service could not internalize the effects of the presence of vacant cabs on the intensity of demand and end up supplying an insufficient service (Bergantino et al., 2007). This possibility is particularly important with reference to the availability of the service in non standard hours, that might be overlooked by the incumbents and that a regulator could guarantee by fixing minimum service standards (basically, by setting compulsory shifts in some specific hours of the day).

Another critical aspect of the service is that fare fixing by a regulator can by itself favour the meeting of demand and supply: in a hypothetical regime of free competition, where, every time the service is needed, the customer and the taxi driver have to bargain bilaterally for the trip price, the results could be very variable, with a high uncertainty for customers and cab drivers and high search costs that could determine an undersizing of the market (Visco Comandini et al., 2004); this supports the usefulness of forms of regulation specifically targeted at increasing transparency, including the setting (to protect consumers) of maximum fares.

Only in the first of the two cases mentioned here the introduction of entry barriers by the regulator could be justified: it would be a compensation for the obligation to supply the service in times of slack demand. Notice that service obligation characterizes taxis but not limousine services (that do not have a service obligation). Some form of registration of all the operators (taxi and NCC) could be anyway necessary in order to

---

4 Clearly the regulator cannot, in the long run, impose the presence of an “excessive” number of operators or shifts in the market (violating their profitability constraint). Therefore in the regulated market the use of the service will be as a rule lower than would have otherwise.
determine who is authorized to use the reserved lanes: the goal to avoid congestion could be achieved through road pricing mechanisms (and the introduction of parking fees) and not necessarily through forms of limitation of entry of new operators (that entering in the market get anyhow a “privilege” in the use of the road network - a fixed resource - and in particular of the reserved lanes).

This last possibility seems to suggest a possible link between the regulation of the service (in terms of rationing of licences and fare fixing) and other public policies concerning traffic and mobility (Beesley, 1973). In fact, the variables $t$, $x$, $p_a$, $w_a$, and $x_a$ are influenced to a large extent by local policies. For example the average length of a trip depends on the dispersion in the area of the main interest points of the city and on the proportion between the fixed and the variable component of the fare; from this last aspect, an increase in the trip average price will increase the demand for longer trips (and therefore $t$) and reduce the shortest ones, if it is achieved mainly through an increase of the fixed component of the fare. Likewise an improvement of mobility conditions for cabs can raise their average speed and reduce $t$ and $w$ (for given $n$ and $p$).

Last, qualitative improvements in transport substitutes can drive down demand for taxi services in favour of other means of transportation and make demand for taxis more elastic (as more easily substitutable) with respect to price and waiting time.

The motivations and the features of the current regulation seem to reflect only partly what indicated so far. The diffused presence of entry barriers seems more linked to forms of protection of the incumbents than to the presence of service obligations: in Italy taxi licences, as well as NCC authorizations, are distributed free of charge, but in limited quantities, in the primary market, having often a very high secondary market value. The restriction in the number of taxis ends up passing on NCC market, given the indirect competition to taxi service that could derive from it.

Other features of Italian regulation do not seem especially geared to efficiency, rather they seem to derive from some forms of pressure on the regulator (Stigler, 1971). For example, the prohibition to cumulate licences or to create taxi companies determines a dispersion of supply which makes it similar the payoff of each incumbent, facilitating the capture of regulator.

Whichever is the origin and the aim of the regulation, it should require, in order to avoid forms of capture of the regulator and to her benefit, the availability of abundant and accurate information on the structural and cyclical situation of the market (Basili, 2008). There is the need of analyses on consumers’ types (residents and visitors, users and non-users, frequent users and occasional ones, business and private users), on their demand elasticity with respect to price, income, and waiting time, on the pattern of services requested in terms of main destinations and average trip length, on actual quality and waiting time, on fixed and variables costs of the provision of the service. Moreover regulation should stem from an evaluation of the role of taxi and NCC in the local public transportation system and from studies on the impact of the different measures of traffic control (parking limitations, reserved lanes, and so on).

Finally, in order to avoid regulatory capture, the level of regulation should be sufficiently “high”. From this point of view the municipal level looks rather vulnerable to capture and not very much representative of customers, given that a big proportion of potential demand is expressed by non residents (Noam, 1982).

---

5 The Bersani law introduced the possibility (de facto not much used) of introducing licensing fees for the new operators (see further on).
3. The national legislation

The Italian legislation on taxi and limousine services (NCC) consists of a national law (general policy law No 21/1992\textsuperscript{6}), regional laws, and municipal bylaws. In 2006 the general policy law has been partially modified by paragraph 6 of law No 248/2006 – Bersani law).\textsuperscript{7}

The changes introduced with the recent legislation try to answer to the widespread protests of customers that complained about the insufficient supply. The need for a reform was strengthened by the poor supply of fast public urban transportation (undergrounds), that tends to increase the use of private cars, with negative effects on road congestion (including parking in prohibited zones) and on average speed of urban public transport by road. The high cost of licences on the secondary market (see section 5) indicated also an excess of demand and the presence of regulation rents. The recent changes in the legislation follow a series of proposals of reform of the sector,\textsuperscript{8} yet they do not modify the general setting. As a matter of fact the law merely offers some additional options to municipalities willing to increase the supply of the service, a power that they already had on the basis of the law of 1992. Moreover some of the options included in the decree initially enacted have been eliminated or substantially rescaled at the moment of turning the decree into a law, following the protests of the taxi drivers; in particular it has been eliminated the possibility for a taxi driver to have more than one licence and the introduction of temporary or seasonal authorizations has been reserved mostly to incumbents.

The national general policy law distinguishes between taxi services, with obligation to park in public spaces and fares regulated by municipalities, and limousine services (NCC), with obligation to park in a garage, non regulated fares, and availability only by phone. Municipalities, on the basis of principles indicated by regional laws, set, by means of specific regulations, the procedures for the service (shifts, and so on), taxi fares, and give taxi licences and authorizations for limousine services.

Licences (authorizations for NCC) are referred to a single vehicle, are given free of charge through an open competition and are transferable. The Bersani law introduces the possibility of assigning licences against payment and in this last case at least 80% of the revenues must be given to the local incumbents. Other ways of increasing supply have also been introduced, by giving municipalities the power to assign (mostly to incumbents) non-transferable specific temporary or seasonal authorizations, allowing incumbents the possibility to use substitutes and additional vehicles for services to specific groups of customers and to provide for innovative services.

Taxi licences are released to individuals (that can be on the artisans’ register, associated in cooperatives or consortia) and cannot be cumulated (as mentioned, this was allowed by the previous decree). Municipalities set the requisites for the granting of the licence (authorization for NCC). It is possible for family members to drive the vehicle instead of the car owner, and in some specific cases (holidays, health problems, and so on) to have other subjects as substitutes in driving. For NCC the same person,

\textsuperscript{6} General policy for the transportation of persons through public car and bus non scheduled services.


\textsuperscript{8} The most important proposals have been made by the Agency for the control and the quality of local public services of the municipality of Roma (AGSPL, 2004a), by the Antitrust authority (AGCM, 1995, 2004) and by several academic contributors (for example, Boitani and Bergantino, 2003; Bordignon and Boitani, 2003).
that can also be a private entrepreneur, can cumulate authorizations and hire workers (and have driver substitutes). The law No 248/2006 attributes to municipalities the faculty to extend the use of driver substitutes for taxi drivers when supplementary shifts are established.

Taxi and NCC services are different also in terms of fares (set by municipalities in the first case, set by operators in the second one) and of mandatory service. For taxis, the law of 2006 has reinforced the power of municipalities to set fares for predetermined trips, an option that could increase the transparency and predictability of customer’s charges, with positive effects on demand.

The national general policy law provides for the creation of an advisory committee (in regions and municipalities), giving it some powers on the service and on the application of the regulation. The Bersani law also provides for the possibility of establishing a committee for the monitoring of regularity, efficiency and adequacy of the service to demand conditions. In both cases the law provides for the participation of representatives of customers and incumbents. The distinction between the functions of the two committees is not clear. Moreover, the competence of the advisory committee on the evaluation of the application of regulations is improper, given the presence of conflicts of interest among the regulator and some members of the committee (regulated). The other functions seem to be basically common to both of them.

In short, the national regulation sets a barrier to entry and the requisite of registration to the specific roll both for taxi and for NCC drivers. Once entered in the market, the two types of operators are subject to different operational constraints: for cabs shifts and fares are established by municipalities, there is obligation to park in specific areas and licences are not cumulable; NCC permit holders can freely decide service hours and fares, must park in garages and licences are cumulable. The decision on the specific content of the regulations (shifts, fares, number of licences, and so on) and the surveillance of their observance is assigned to municipalities, that are supported by two committees, an advisory committee (provided by the national law, but with very vague tasks), and a monitoring committee, with better defined tasks, but whose creation is optional.

The types of regulation of the sector at the international level are very varied (OECD, 2007). With reference to the taxi market, in some countries (Sweden, Netherlands, Ireland, Austria, Hungary, Australia, and New Zealand) a liberalization process started in the nineties and mainly abolished quantitative entry barriers of operators. The reductions of the constraints to the societal organization and the liberalization of fares have been less frequent.

Even in the countries where the sector has been more liberalized, no significant market failures have been reported. Sweden and New Zealand started a broad liberalization including tariffs, but keeping some standards on service quality. In Norway, in the areas where there are several competing radio-dispatching firms, there has been a pricing deregulation with a continuation of entry quantity barriers. Local antitrust authority found a fare increase after deregulation, especially in hours and days with low demand. In London new drivers have to take a special driving test and a street knowledge test. In the case of Ireland, the effect of the drastic reduction in the value of the licences after market liberalization induced the regulator to devise forms of monetary compensation for some groups of operators (in other words, it is a

---

9 Roll of drivers of vehicles or water-crafts used as public non scheduled transport services established in the local Chamber of commerce.
compensation for the cancellation of a rent created by the regulator itself). Three years after the liberalization of taxi licences in Ireland it has been created the Commission for taxi regulation, that centralizes at the national level the regulation of the sector. The Commission also carries out surveys on the different types of customers and operators, so as to increase the efficiency and the quality of the service. In 2005 the Commission introduced a single national maximum fare, on the basis of a study that showed the absence of differences in the local costs of the service. In Dublin the elimination of the administrative barriers to entry determined the tripling of circulating taxis in two years without significant deterioration in service quality (Barrett, 2003; Daly, 2004). A similar effect on supply has been registered in New Zealand.

In France, a country that had so far a quite strict regulation, at the beginning of 2008 the Attali Commission proposed to give a taxi licence free of charge to anyone who had applied for it up to the end of 2007 and did not get it (thus adjusting effective supply to the potential one) and to authorize the use of the same car/licence by more than one driver (subjecting it only to the observance of a maximum number of hours for each driver, set by law and monitored with devices installed in the car). In municipalities where the evolution in taxi availability did not follow population trends, the prefects could take the mayor’s place in the management of taxi licences. At the national level the Ministry of Transport would regulate VPRs (similar to Italian NCC), while in the case of taxis it would be responsible for regulation together with the Ministry of the Interior, so as to avoid the concentration of regulation of the two segments of the market under the same ministry. Despite the renunciation by the French government of the Attali plan after the protests of the taxi drivers, at the end of May 2008 is has been accepted the proposal of the minister of the Interior that provides, among other things, for: i) an increase of more than 25% in cabs circulating in Paris (from 15,600 to 20,000) before the end of 2010, of which 1,200 within 2008; ii) greater transparency in receipts, that will contain details of all fare components applied; iii) a reserved lane on the highway to the main airport.  

4. Regional and local regulation

In the first months of 2007 the Bank of Italy conducted a survey on taxi and NCC service in 67 Italian municipalities, localized in all Italian regions, with more than 50,000 inhabitants at the end of 2005. One of them did not participate to the survey; in some other cases the questionnaire has been returned incomplete. The cities have been chosen also on the basis of their relevance as a tourist attraction; in the case of Venezia

---

10 In mid 2008 the cost of a taxi licence in Paris was about € 180,000; many drivers rented a car for about € 700 per week, and at the end of 2007 the tax abatement on fuel reserved to taxi drivers was abolished.

11 The cities where the survey was conducted are: Alessandria, Ancona, Aosta, Arezzo, Asti, Bari, Bergamo, Bologna, Bolzano, Brescia, Brindisi, Cagliari, Campobasso, Caserta, Catania, Catanzaro, Como, Cosenza, Cremona, Ferrara, Firenze, Foggia, Forlì, Genova, Grosseto, La Spezia, L’Aquila, Latina, Lecce, Livorno, Lucca, Messina, Milano, Modena, Napoli, Novara, Padova, Palermo, Parma, Pavia, Perugia, Pesaro, Pescara, Piacenza, Pisa, Pistoia, Potenza, Ravenna, Reggio di Calabria, Reggio nell’Emilia, Rimini, Roma, Salerno, Sassari, Siena, Siracusa, Taranto, Terni, Torino, Trento, Treviso, Trieste, Udine, Varese, Venezia, Verona, Vicenza. Rimini did not return the questionnaire in time. Almost half of the administrations interviewed did not give information on the dates of the second-last assignment of taxi licences and almost 40% on the second-last fare change.
the survey concerned both the land and the lagoon services. Cities have been divided in three groups by population size. Big cities include 5 towns with more than 500,000 inhabitants (Roma, Milano, Napoli, Palermo, Genova). Middle-size cities are 7 and include municipalities with a population between 250,000 and 500,000 units. The survey focused both on institutional aspects (norms, regulatory bodies, extent of agreement procedures among the interested parties, degree of decentralization in decisions) and on the market impact of the regulation in terms of licence number, level and fare structure, and finally on the variables considered by local regulators in order to decide a change in fares or in the number of licences. In this section the results of the survey on the institutional aspects are presented; the next section focuses on market situation and dynamics. The information asked in the questionnaire refer to the situation at the end of 2006.

Beside the national legislation and the municipal regulations, recently Regions have legislated on the matter, due to the new powers given to them by the legislative decree 422/1997 and by the reform of title V of Constitution.\textsuperscript{12}

In two third of the cases the Regions issued a specific law, 19\% included the matter within the local public transportation regulation (and/or the regional transportation plan), in the remaining cases Regions did not legislate on the subject (Table 1). The Molise Region neither has a regional legislation nor a specific roll in the local Chambers of commerce. In 4 municipalities over 66 there are no local regulations for the sector and 10 municipalities did not answer to the specific question. In one case the Region and in two cases the Province have the power to approve the municipal regulations.

In many cases the local legislation (both regional and municipal) repeats widely and follows closely the national one. The setting of the number of licences (and of NCC authorizations), of the level of fares and their change is left usually to the municipalities, often on the basis of general criteria (see further on). Some Regions, mainly those with important airport basins (in particular Lombardia) have also super-municipal regulations.

\textsuperscript{12} The new article No 117 of Constitution does not mention local public transport neither among the matters of exclusive state competence nor among those of concurrent legislation. In this context the Constitutional Court (see decision No 222 of 2005) held that the regulation of this sector falls under the residual competences of the Regions specified in paragraph 4 of the article 117 of the Constitution. Moreover the Constitutional Court has rejected the appeal made by the Veneto Region against article 6 of the law No 248/2006 (concerning taxis) because the norms impugned do not discipline the organizational procedures and the carrying out of the service, but are referable to the cross-competencies in the matter of defence of competition on which the State has exclusive legislative power (see decision No 452 of 2007).
Table 1: Features of local regulation on taxi and NCC (percentages).

<table>
<thead>
<tr>
<th>REGIONAL LAW</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, specific</td>
<td>Yes, included in the law on local public transportation</td>
</tr>
<tr>
<td><strong>66.7</strong></td>
<td><strong>19.0</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REGIONAL ADVISORY COMMITTEE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>38.1</strong></td>
<td><strong>61.9</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROVINCIAL ADVISORY COMMITTEES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>9.5</strong></td>
<td><strong>90.5</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPROVAL OF MUNICIPAL REGULATIONS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Only the Municipality</td>
<td>The Province too</td>
</tr>
<tr>
<td><strong>19.0</strong></td>
<td><strong>9.5</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUPER-MUNICIPAL REGULATIONS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>14.3</strong></td>
<td><strong>85.7</strong></td>
</tr>
</tbody>
</table>

Note: The provinces of Trento and Bolzano have been considered separately.

More than one third of the Regions established advisory committees at the regional level and 9.5% also at the provincial level. 30.8% of the municipalities interviewed did not establish the advisory committee. In the municipalities where it is present, the trade associations (taxi, NCC, radio-dispatching companies) prevail among its members (on average they are more than one third of the total), while the representatives of consumers’ associations are a minority (on average less than one sixth of the total; Figure 1).

The high representation of the category reflects the willingness of the local administrations to allow the participation of the many trade unions and trade associations (artisans, cooperatives), of which the industry operators are members. Moreover, the wide range, in terms of political affiliation, of taxi, NCC and radio-dispatching associations present in each city seems to reflect the need of the industry to “capture” the regulator, irrespectively of the political group in power locally at each point in time. Among the main cities, the ratio between representatives of the category/customers’ representatives is particularly unbalanced in Torino and Genova (13/1 and 11/1, respectively). In 5 cities consumers are not represented in the committee. Only one municipality does not include in the committee neither operators’ nor consumers’ representatives.
Figure 1: Composition of the taxi advisory committee in the main Italian cities (percentages).

These data can be summarized with a “capture” index, given by the ratio between the number of the operators’ representatives and the sum of the last variable and the number of consumers’ representatives. The index varies between 0 and 100 (maximum capture). Using our dataset it takes a mean value of 68.8 and it decreases with city size: is it equal to 64.9 in small cities, to 74.2 in middle size cities and to 86.7 in big cities.

Only 4 municipalities have appointed the monitoring committee indicated in the Bersani law. With the exception of Pavia, with 2 consumers’ representatives and 1 delegate of the operators, in these committees customers are a minority too. The new municipal regulation of Roma (presented at the end of 2006, revised a year after, not yet approved in April 2009) establishes an advisory committee with 3 experts of the industry, 5 delegates of the category (3 for taxi drivers and 2 for NCC), and 3 consumers’ delegates. In the previous committee there were 17 delegates of the operators and 2 of consumers. It is also established that the town council should present to the governing council a yearly report on the monitoring of the sector.

Local regulations do not mention, with rare exceptions, the use of indicators of demand in the evaluation of the adequacy of the number of taxi and NCC. In the case of the Piemonte region, the regional law gives to Provinces the task to indicate to the municipalities “measures of restraint of licences and of authorizations” on the basis of: population, territory size and other specific characteristics, size of tourism and business flows, relative supply of other local public transportation, other salient factors for local transportation of persons. The Province of Novara, for example, elaborated specific formulas of congruity of taxi and NCC supply. The law of Umbria Region specifies that taxi and NCC numbers must be decided by municipalities on the basis of population size and on the relevance of tourism, business, social, and cultural activities in the cities and in the surroundings, on the distances among urban poles (railway station, centre-periphery, and so on), and on supply of local scheduled urban and suburban public transportation. The regional laws of Valle d’Aosta and Veneto specify that the regional council sets the maximum number of licences and/or authorizations for each area on the basis of indicators. The Veneto Region requires that at least 5% of licences is given for the transportation of disabled persons. In Lazio and Emilia-Romagna the setting of the criteria is attributed to the Provinces.
In most cases the more recent decisions to assign taxi licences and/or NCC authorizations have not been taken on the basis of market indicators (68.7% of interviewed municipalities). A small share of municipalities answered they took into consideration generically demand (10.4%) and a similar share of municipalities specified instead the group of indicators used. A 2003 study of the province of Parma estimated the effective need for taxi and NCC on the basis of the average ratio of service supply/population of similar Italian provinces (and detected a shortage). In Parma it has also been estimated future yearly demand on the assumption that its average increase would be similar to that of total demand for mobility in the region Emilia-Romagna (3%), that should concentrate on public transportation. In the same year the Region Lombardia presented an analysis comparing demand and supply in the airport basin and in the Milano area that points out the need of a different articulation of shifts and of a limited increase in the number of licences. In the survey of Banca d’Italia only in one case the new licence issue has been based on a specific request of the operators.

5. Situation and dynamics of the market

The analytical scheme of section 2 shows that the main variables describing a taxi market are price (more in general the fare structure), the level and the distribution of effective supply in the area and in the daytime, the demand characteristics and in particular its relationship with the waiting time function. A second group of relevant variables are potential demand and unsatisfied demand (that give information on the height of barriers to entry), potential supply (that gives information on rents), and service costs. A third group includes the context variables (extension and characteristics of the territory and of the urbanization), the availability of strict substitutes for public use (km of underground) and local policies already implemented that can influence the efficiency of the service (length and degree of traffic restrictions in the traffic limited zone, reserved lanes, and so on).

With reference to this last set of variables, the sample of the cities included in the survey of Banca d’Italia is quite diversified. The size of the land area of each city is very different even inside the same class-size of the cities. Roma is seven times wider than Milano; Ravenna is the second Italian city in terms of square kilometres. In some small cities (Bergamo, Siena, Pisa, and Pavia) the traffic limited zone is wider than in Firenze. Among the first 10 cities in terms of population 5 have an underground network (Roma, Milano, Napoli, Genova, and Torino). Among the big and middle-size cities in the sample, Firenze and Bologna have the wider traffic limited zone in km per inhabitant (10.11 and 8.43 square metres per inhabitant, respectively), followed by a second group (Palermo, Napoli, Verona, and Roma, with 5.66, 3.49, 3.43, and 2.44 square metres/inhabitant).^13^ In the following sections there will be presented some information on the two group of variables.

^13^ The data quoted here are taken from Istat.
5.1. The number of taxi licences and NCC authorizations

The results of the Banca d’Italia survey indicate that the average number of taxis per 10,000 inhabitants is 20.8 for the big cities, 12.2 for the middle-size ones, and 3.4 for the small ones (Table 2). For NCC the ratio with respect to population is on average 2.3, with a more homogeneous distribution with respect to that of taxi. The variance is quite high also inside each group of cities (Figure 2).

Table 2: Distribution of taxi and NCC supply in the main Italian cities.

<table>
<thead>
<tr>
<th>Type of city</th>
<th>Taxi licences per 10,000 inhabitants</th>
<th>NCC authorizations per 10,000 inhabitants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Minimum</td>
</tr>
<tr>
<td>Small cities</td>
<td>3.4</td>
<td>0.7</td>
</tr>
<tr>
<td>Middle-size cities</td>
<td>12.2</td>
<td>3.9</td>
</tr>
<tr>
<td>Big cities</td>
<td>20.8</td>
<td>4.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5.8</strong></td>
<td><strong>0.7</strong></td>
</tr>
</tbody>
</table>


In general taxi availability is weakly correlated with NCC supply (the correlation coefficient is equal to 0.24), and there are on average 7 NCC available every 10 taxis. Yet the ratio between NCC and taxi tends to be inversely related with city size (from 0.2 for big cities to 0.5 for middle-size cities to 0.8 for small ones\(^{14}\)), and it is higher in tourist centres where demand fluctuates seasonally. This plausibly reflects the fact that in these cases it would be costlier to set up a service with specific obligations (shifts, and so on).

Relating licences to resident population allows to scale the data of the different cities but does not give precise information on service adequacy with respect to potential demand. An effective regulation, that does not want to take the risk of restricting supply excessively, should be based on a careful examination of demand conditions and of service profitability in the different hours and periods. In Italy the surveys done to this end are very rare.

In Roma two surveys on residents have been conducted. The first one showed an average unsatisfied demand (because the interviewed person could not find a vacant taxi) of 20%, with peaks up to 27% in December, waiting time from 4 to 24 minutes (on average 15 minutes for the suburbs) and a ratio of non-dispatched demand by the radio-dispatching centres of 20-30% (Sta, 2001). The second, conducted by the Agency for the control and the quality of local public services of Roma (AGSPL) in 2003, found that: i) 18% of residents use taxis (on average one trip every two months); this means that more than 80% of interviewed people never or hardly ever use taxis; ii) more than 8% does not use taxis (or does not use them more frequently) because it does not save enough time, more than one third because taxis are too costly, 10% and 5.4%, respectively (among those that use them in the area of residence or in the area of work) because it is difficult to find a vacant one (AGSPL, 2004b, 2004c). A survey done by ACI-Eurispes (2006) finds that 1.2% of interviewed people use taxis often or

\(^{14}\) Among these cities, the 6 with a ratio NCC/taxi particularly high are: Potenza (5), Perugia, and Ravenna (2.2), Campobasso (1.7), Novara (1.6), Bergamo (1.4).
systematically and 79.2% never turns to them. A survey by Altroconsumo (2007), conducted in the last months of 2006, finds a particularly long waiting time at the main railways stations’ parkings and at the airports of Bologna, Firenze, Milano, and Roma.

In Table 3 we tried to supplement the simple ratio of licences to population in the municipality with other indicators that could proxy potential demand such as the number of firms, tourists, occasional or frequent visitors of the city for study or business purposes, the territorial extension of the territory, the presence in the surroundings of an airport or of trade fair facilities, the subway extension, and so on. For simplicity, we consider only the 11 greatest Italian cities (in terms of population).

The table shows that often the ordering changes with the indicator considered. For example for Milano the index based on municipality population exceeds that of Roma by more than 50%, but it is very similar to the latter when the provincial population is considered, and reduces further if the provincial firms are taken. The relative position of Bologna moves down from the fifth to the tenth position if instead of population (provincial or of the municipality) one takes the firms number. In the case of the ratio licences/tourists Roma falls to the fifth place, after Milano, Torino, Napoli, and Genova. The capital looks relatively more deficient of taxis also if one consider as a scale
variable the number of airport passengers, falling at the fifth place, and Milano falling to the seventh one.

Table 3: Distribution of taxi supply in the main Italian cities with respect to several reference parameters (indices: Roma=100).

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Population resident in the municipality</th>
<th>Population resident in the province</th>
<th>Licences/population of the province</th>
<th>Licences/ firms in the province</th>
<th>Licences/ tourists' arrivals</th>
<th>Licences/ airport's passengers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roma</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Milano</td>
<td>51.4</td>
<td>96.8</td>
<td>155.6</td>
<td>82.6</td>
<td>52.9</td>
<td>139.6</td>
</tr>
<tr>
<td>Napoli</td>
<td>38.6</td>
<td>76.8</td>
<td>101.1</td>
<td>50.8</td>
<td>179.0</td>
<td>123.6</td>
</tr>
<tr>
<td>Palermo</td>
<td>26.3</td>
<td>30.9</td>
<td>20.0</td>
<td>17.0</td>
<td>96.7</td>
<td>39.9</td>
</tr>
<tr>
<td>Genova</td>
<td>24.3</td>
<td>22.1</td>
<td>58.8</td>
<td>64.7</td>
<td>111.0</td>
<td>106.0</td>
</tr>
<tr>
<td>Torino</td>
<td>17.0</td>
<td>56.0</td>
<td>147.1</td>
<td>44.7</td>
<td>51.2</td>
<td>137.3</td>
</tr>
<tr>
<td>Bologna</td>
<td>14.7</td>
<td>23.8</td>
<td>73.7</td>
<td>45.5</td>
<td>45.2</td>
<td>68.4</td>
</tr>
<tr>
<td>Firenze</td>
<td>14.4</td>
<td>24.2</td>
<td>67.9</td>
<td>40.4</td>
<td>56.8</td>
<td>23.0</td>
</tr>
<tr>
<td>Bari</td>
<td>12.8</td>
<td>39.8</td>
<td>19.6</td>
<td>6.3</td>
<td>17.6</td>
<td>37.1</td>
</tr>
<tr>
<td>Catania</td>
<td>11.9</td>
<td>26.8</td>
<td>26.1</td>
<td>11.6</td>
<td>63.8</td>
<td>39.6</td>
</tr>
<tr>
<td>Venezia</td>
<td>10.6</td>
<td>20.8</td>
<td>16.2</td>
<td>2.2</td>
<td>98.9</td>
<td>8.2</td>
</tr>
</tbody>
</table>


Note: Data on population resident in the municipality refer to the end of 2005, data on population resident in the province refer to the end of 2006. Data on firms refer to 2004 (database Asia), tourists refer to the province and to 2005, airports’ passengers data refer to 2006 and exclude transits. For Venezia only land licences are considered.

In dynamic terms these data show that, while population was basically static in the last ten years, almost always the other indicators of potential demand had an expansionary trend. For example, between 1991 and 2006 passengers’ traffic at the airports more than doubled in Roma, almost tripled in Milano (exceeding 30 mln in both cases), more than tripled in Bologna. In Venezia the increase has been greater than 300%; even higher growth rates have been registered in some minor airports, among which Bergamo (from almost 250,000 to more than 5.2 mln), Treviso (from more than 84,000 to 1.3 mln), and Forlì. From 1990 to 2005 tourists’ arrivals in the respective provinces tripled in Milano, Venezia, and Roma (in this last city they went from 5.5 to 17.4 mln), in Bologna they increased from a little more than 1 mln to 2.7 mln. Last, in 2000-04 the number of visitors to Italian trade fairs increased on average by 11.5% (Censis, 2006).

Against these phenomena, the adjustment in the supply of taxi and NCC has been fragmentary and very limited. 41.8% of the municipalities in the sample assigned the last licences more than 20 years ago (43.3% for NCC; Figure 3).

For taxis 12 municipalities never had selections for the assignment of new licences or last selections date back to the sixties; 6 municipalities did not give any answer to this question. In the 5 years preceding the survey there is a resumption of taxi licences assignment for about one third of the municipalities. A similar trend is found for NCC, for which a little more than two fifth of the municipalities issued the last authorizations more than 20 years ago and a similar share did that in the last 5 years (no authorizations have been issued from 16 to 20 years ago). In the more recent years (before the Bersani...
law) it seems that municipalities tried to increase the service mostly through new NCC authorizations. Yet it does not emerge a clear relationship between the recent issue of licences/authorizations and relative shortage (in terms of the ranking of the city by level of NCC authorizations previous to the new issue). Among the municipalities that issued them more recently, only 3 had more taxis than the average.

The criteria chosen by municipalities to rank applicants for licences/authorizations assignment are of various kind. Among the preferential qualifications of applicants, very often the more relevant one is to have been a substitute driver for a certain amount
of time. In this way family members of the drivers are favoured, given that they have higher probability of having done this for some time. Generally in the main cities and in those preferred by tourists the knowledge of a foreign language is positively valued. The choice of other preferential qualifications goes from the educational degree (this is the case of Roma municipality in 2001 selection) to the registration in the unemployment or mobility rolls (regulation of the Reggio nell’Emilia municipality) and includes sometimes indicators of the dependency burden.

After the Bersani law there have been no generalized changes in the industry. Yet the main cities implemented some relevant initiatives. In some of them the monitoring committee has been introduced. Among the most interesting innovations, in Roma there has been an increase in the number and in the length of shifts, in the cases where it is possible to use substitutes in taxi driving. As a whole these measures should have increased the effective service in peak times of about 2,500 taxis. The municipality has also assigned to Censis research institute a study on the evaluation of taxi necessity in town and to ATAC spa a feasibility study of a technologically advanced monitoring system. After the taxi drivers representatives showed their opposition to the use of any satellite control system, the municipality issued (free of charge) about 1,950 licences (until the first half of 2008) and intended to issue 250 more before June 2009. When all this project will be concluded, the number of licences should be higher by 30%. In Milan the increase in the service supply has been obtained by a reorganization of shifts, also with the option of drivers to hire drive substitutes. Service monitoring has been done through remote connections with the radio-dispatching centres. In Firenze 60 new licences have been issued together with 30 temporary ones (+15.2%), the shift time has been increased and new types of service have been introduced. In Bologna, were no new licences have been issued since 1983, in March 2007 there has been an agreement of the municipality with taxi representatives to issue (for a fee) 42 licences (+6.4%, from an initial proposal of about 130), part of which will be used for preferential service for disabled people. In June 2008 the municipality opened a competition for 41 licences, of which 23 for disabled preferential service (sold for € 150,000) and 18 with some restrictions on the service (type and area of demand, sold for € 125,000), and a rebate of € 12,500 in the case of the use of a hybrid and/or electric vehicle. 90 people applied for the licences. Licences have been assigned in March 2009 (two years after the agreement). In Modena an agreement between taxi drivers and the municipality (April 2008) provides for the issue of 10 licences (for a fee) and 5 temporary licences together with a fare increase of 8.5%. In Perugia the municipality has planned to issue 8 new licences (+28.6%), while in Siena it has been introduced the possibility to have more shifts per car by hiring driver substitutes; in the mid of 2007 a selection started for 2 taxi licences and 3 NCC authorizations. In Napoli the municipality carried out a survey to evaluate taxi availability in taxi ranks in the different times of the day. In Pisa in 2007 12 new licences have been issued (+25%). In Palermo 80 licences were already assigned in 2006 (+30%), after a selection started in 2005.

Beside the supply increase, the municipalities of Roma and Milano implemented some policies to improve the service, among which the creation of a single telephone number for calling taxis at their ranks and some measures to contrast unauthorized

15 The cost of licences has been set at a level lower than the secondary market price given the constraints associated to these new licences. If demand is greater than supply the municipality selects applicants by testing them on their knowledge of the area (roads, cultural sites and so on), of a foreign language and by assigning some points to the school degree attained and to the status of unemployed.
service. In Milan the agreement of November 2006 with the operators associations also provides some measures to improve taxi mobility (increase of reserved lanes, enforcement of controls for the no parking and restricted traffic areas, activation of new cameras). Moreover the municipality introduced subsidies for the renewal of the car pool, of radio equipment, the installation of POS facilities, the use of satellite navigator systems, and so on (Ministero dello sviluppo economico, 2007).

Other indicators of supply adequacy can be derived from the presence of entry queues and of rents arising from the possession of a licence, licence that is primarily distributed for free. From the first point of view, an indication can be drawn from the number of subjects with the minimum requisites to participate to a licence issuing session (in Italy the registration to the specific roll established in the provincial Chambers of commerce) and from number of applications to the public contests for a licence.

Figure 4: Ratio of taxi and NCC over the number of persons registered in the rolls of drivers for non scheduled transport services.

The survey shows that in the main Italian cities for each licence holder there is at least another subject registered in the roll established in the provincial Chamber of commerce that could, potentially, participate to the municipal contests to allocate licences or NCC authorizations (Figure 4).\(^\text{16}\)

Information on the price of licences on the secondary market are scarce and sporadic. The only one done systematically has been carried out in Roma in 2003 by submitting a questionnaire to a sample of operators, and asked for information on the period 1980-2002. The survey shows that, compared to an average investment in real estate in Roma municipality, from 1986 to 2002 the purchase of a taxi licence would have had a higher yield by about 0.5 percentage points (Visco Comandini et al., 2004). The recent value of licences varies, according to various sources, from € 50-60,000 in Bari to € 300,000 in

\(^{16}\) This indicator has a limited informative capacity because it is influenced by the licences availability on the secondary market, by the diffusion of driver substitutes (registered in the rolls too) and by the time proximity of new contests (in that case normally there is an increase in registrations). This information has been given very partially by the municipalities interviewed: 12% of them did not give any information, in 3 regions the roll has not been created and in Friuli-Venezia Giulia the roll is regional instead of provincial.
Firenze; according to a survey of the municipality, in Bologna the licences value is around € 250,000 for taxis, and € 125,000 for the NCC.\textsuperscript{17}

The actual supply of taxis does not depend only on the number of licences, but also on the possibility and the actual use of the same vehicle by drivers’ substitutes, on the number of shifts and hour ceilings fixed by local regulation, and on the actual schedule of individual taxi drivers. The number of service hours guaranteed in a typical working day (vehicle-hour) can therefore vary considerably with respect to the number of licences too, according to municipal regulations. If, as in the Italian case, one licence corresponds to one vehicle (but it is possible in some case the use of driver’s substitutes), the potential maximum supply is given by having all vehicles circulating 24 hours a day. With respect to this potential ceiling, the regulation sets compulsory and optional shifts, maximum working schedules for the individual taxi driver (for example 9.2 hours in Napoli, 10 hours in Roma, 12 hours in Torino, Genova, Bologna, and Firenze) and limits to the possibility of using driver’s substitutes.

It is likely that if in this type of choices operators have a wide margin of action, they would choose a schedule that concentrates taxi availability in the hours of higher demand, at the expense of non-peak hours. The Banca d’Italia survey indicates that for about one third of municipalities shifts are organized by the operators’ organizations and in a little more than 12\% of the cities shifts are not regulated; in Palermo there are regulated shifts only for the airport service.

Another aspect which is relevant for the actual availability of the service is the choice of drivers to work in the optional shifts, and the possibility for the municipality to monitor the actual service provision in the compulsory shifts and the compliance with the hours ceiling. While it is basically impossible to verify the actual work of the operator in the established shifts (in the absence of remote control mechanisms on the use of the vehicle), it is quite simple to check the extra work. The actual supply is therefore undetermined to a great degree, and, where they have been set, the hour ceilings for each shift determine the maximum actual supply. Moreover, given the different route profitability, there could be some cream skimming effects on some routes, with a territorial concentration of supply.

The actual supply of the service depends also on the possibility, given by some regional regulation, to pick up the client also outside the territory of the municipality that issued the licence. Apart from specific exceptions when there is an airport in the outskirts (normally in these cases the service is regulated by agreements among the interested municipalities), the survey showed that this possibility is provided in 10.4\% and 19.4\% of the municipalities for taxi and NCC, respectively.

5.2. Taxi fares

Fares for a 5 km trip on a working day (and working time) vary from 3.7 to 15.2 euro, and the average is 8 euro (Table 4). These data do not take into account the time charge that applies when the speed is lower than a certain threshold (normally 20 km/hour), for example at the traffic lights or in the case of traffic jam. The results of the survey show a high fare variance, especially in the fixed component and in particular for middle-size cities.

Table 4: Structure of urban taxi fare in the main Italian cities by city size (€).

<table>
<thead>
<tr>
<th>Type of city</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Difference between min. and max.</th>
<th>Coefficient of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FIXED COMPONENT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small cities</td>
<td>3.0</td>
<td>1.7</td>
<td>5.2</td>
<td>3.5</td>
<td>0.22</td>
</tr>
<tr>
<td>Middle-size cities</td>
<td>3.7</td>
<td>1.8</td>
<td>8.7</td>
<td>6.9</td>
<td>0.57</td>
</tr>
<tr>
<td>Big cities</td>
<td>2.7</td>
<td>2.3</td>
<td>3.3</td>
<td>0.9</td>
<td>0.12</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3.0</strong></td>
<td><strong>1.7</strong></td>
<td><strong>8.7</strong></td>
<td><strong>7.0</strong></td>
<td><strong>0.32</strong></td>
</tr>
<tr>
<td><strong>VARIABLE COMPONENT (PER KM)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small cities</td>
<td>1.0</td>
<td>0.1</td>
<td>1.6</td>
<td>1.5</td>
<td>0.28</td>
</tr>
<tr>
<td>Middle-size cities</td>
<td>0.9</td>
<td>0.7</td>
<td>1.3</td>
<td>0.6</td>
<td>0.23</td>
</tr>
<tr>
<td>Big cities</td>
<td>0.8</td>
<td>0.7</td>
<td>0.9</td>
<td>0.2</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1.0</strong></td>
<td><strong>0.1</strong></td>
<td><strong>1.6</strong></td>
<td><strong>1.5</strong></td>
<td><strong>0.27</strong></td>
</tr>
<tr>
<td><strong>TOTAL COST FOR 5 KM TRIP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small cities</td>
<td>8.0</td>
<td>3.7</td>
<td>11.0</td>
<td>7.4</td>
<td>0.18</td>
</tr>
<tr>
<td>Middle-size cities</td>
<td>8.4</td>
<td>5.9</td>
<td>15.2</td>
<td>9.3</td>
<td>0.33</td>
</tr>
<tr>
<td>Big cities</td>
<td>6.8</td>
<td>6.2</td>
<td>7.8</td>
<td>1.6</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8.0</strong></td>
<td><strong>3.7</strong></td>
<td><strong>15.2</strong></td>
<td><strong>11.6</strong></td>
<td><strong>0.21</strong></td>
</tr>
</tbody>
</table>


(1) Data refer to the end of 2006. Weekday urban fares in business hours.

Fares are lower in the big cities than elsewhere. A possible explanation of this difference is that it reflects the bigger market size and the presence of concentration poles of users, such as airports and main railway stations, that tend to reduce waiting time of customers and empty trips. The greater congestion of big cities also tends to increase the actual cost vis-à-vis the theoretical one more than in the other urban centres; for example, with the fare in force in Roma in 2007, a 5 minutes car stop (at the traffic lights or when the car slows down in the traffic jam, the time-fare applies) in a 5 km trip increases the total cost of about one third.

A meaningful comparison of service cost in the different cities should indeed take into account, together with the theoretical cost for a given trip, its significance, that is its closeness to the actual consumer’s expenditure. The latter depends on congestion as well as the number of supplements and of fares different from the basic one, that reduce the fare transparency and leave room for operators’ opportunistic behaviour. Among supplements, those that affect more the final price are the booking charges that are levied on customers who book taxis by telephone (included the dead-running cost) and the luggage charges. For example in Roma in 2007 the working day fare for a 5 km trip with a 5 minute time charge (for stops) and a luggage was € 8.99, more than 44% the basic one (€ 6.23). The survey of Banca d’Italia indicates an average of 4.5 supplements in the fare book.

In the determination of the final cost for the user it is also possible some interference of the taxi driver that can choose a trip longer than necessary, arriving at the place of the

---

18 In this computation we did not consider the value of the single fare unit, but simply the hypothetical time and km fares. At the beginning of 2008 in Roma the supplement for the first piece of luggage has been eliminated (see further on).
call with a taximeter already very high, or considering as a luggage a bag which is actually smaller than the regulated limit. In the survey done by Altroconsumo (2007) 44% of taxis arrived at the place were the booking was made more than 5 minutes earlier (the time charges apply) or later than agreed; in Firenze the amount indicated in the taximeter at destination varied from 2.74 to 9.64 euro. In 9% of the cases the initial charge was higher than expected or it has been increased by applying undue supplements. This type of effects could be limited by introducing receipts more detailed and with fiscal effect, that could increase transparency in the customer’s relationship. The new taxi regulation of Roma, not yet approved in April 2009, provides that taxis should have a device connected to the taximeter that can, on request, issue a receipt with the trip details (fares, trip length, supplements applied, and so on). In Bologna the regulation sets a ceiling on the price indicated in the taximeter at the arrival of the taxi at the calling place (a similar ceiling was set by the new regulation of Roma). Another instrument to increase transparency and reduce the opportunistic behaviour of drivers is the introduction of flat rates. In 20.9% of municipalities there are flat rates, generally for the airport (for example in Roma and Milano); in Torino it has been introduced a flat rate inside the traffic-limited zone, in Napoli flat rates for several routes have been introduced.

The use of a two-part tariff is generalized for taxis, with higher fares both for the fixed and for the variable part in days and times of soft demand. In the smaller cities group it is higher both the average fixed part than the fare per km. The economic theory suggests the use of a two-part tariff (as an alternative to average cost pricing) for regulated natural monopolies, so as to allow the firm to recover high fixed costs. A part from the level of fixed costs another motivation for using a two part tariff for taxis is given by the presence of an opportunity cost for trips, because when the driver takes on a customer she gives up the chance of finding a more profitable one (by increasing search and waiting time costs). In this sense the adoption of a fare with a fixed part should provide incentives to avoid an opportunistic behaviour (both unlawfully or simply by operating only in areas where longer trips are requested, such as airports). Moreover the different level of the two fare components could reflect the aim to guarantee a similar average income to the operators in cities with different length/average distances. For the first 8 Italian cities there are no significant differences in the contribution of the two tariff components to the total cost (Figure 5).

On the whole, there is no evidence of a relationship between taxi supply and fare level (Figure 6). In more than 50% of the sample fare changes have been done at intervals at most of three years. This share includes mostly the cities of Centre and North of Italy, while in the South and in Islands fare adjustments tend to be done at larger time intervals. Fare adjustments are never decided on the basis of demand indicators (analysis of elasticity, of demand intensity in “soft” hours, and so on), but simply tend to follow closely the change of the Istat index of consumer prices; in some cases the specific input costs are also considered (car insurance, fuel cost, and so on).

Data on local cost differentials are not available. Some methodological indications can be drawn, with some caution, by a study by the Irish Commission for taxi regulation (2005), that for Ireland does not find significant territorial differences, except for fuel and insurance costs. Average employee labour incomes show instead some variance

---

19 This may happen, for example, if the driver inserts in the taximeter the night or Sunday starting fare, which is higher, when another fare applies.
around the national mean, that comes down considerably when considering industries separately.

For Italy, the ACI indicator of operating cost/km gives some aggregated information. For a diesel car\textsuperscript{20} cost per km vary between 0,30 and 0,29 euro respectively for an annual run of 50,000 and 60,000 km. Fuel weighs 32.7\% upon total variable costs and maintenance has a weight only marginally lower.

![Figure 5: Contribution of the fixed component and of the distance charge to the total taxi cost for a urban journey length of 5 km in the main Italian cities (percentages). Source: survey of Banca d’Italia on the supply and price of taxi service, 2007. Note: Data refer to the end of 2006. Weekday urban fares in business hours.]

Among fixed costs, which represent between 23.2\% and 20.1\% of total costs, car CDW insurance is the most significant one (58.7\%). Another factor that influences the cost function is the use, established by the national regulation (with some specific exceptions introduced by the Bersani law), of the taxi in only one shift, that hampers the distribution of fixed costs on a higher number of km. Radio-dispatching services\textsuperscript{21} are estimated to cost about 170 euro per month by Asso Taxi (and normally they are compensated both by a specific fixed component of the fare and by the fact that taxi drivers set the taximeter on at the moment of the customer’s call and until they reach their client\textsuperscript{22}). There are also additional costs (with respect to the ACI indicator) of CDW insurance premium, due to the higher frequency of accidents of taxis. According to ANIA data, in 2005 this frequency was on average 21.98\% against 7.17\% for the other cars\textsuperscript{23}. Service costs are also influenced by a partial refund of oil excises. Last, the

\textsuperscript{20} Data refer to a car with a power of 1,500-2,000 and to March 2007.

\textsuperscript{21} According to the sector tax benchmarking study, the share of taxi drivers with a radio-dispatching service is about 32\% in the low population density areas and about 90\% and 78\% respectively in the medium/high and high population density areas.

\textsuperscript{22} Moreover almost all radio-taxi firms are cooperatives whose members are the same taxi drivers that make use of the radio-taxi services and that participate to the radio-taxi company profits.

\textsuperscript{23} Data are very variable across cities and oscillate (for taxis) from 5\% in Ferrara to 61\% in Caserta. For the big cities the taxi frequency of accidents varies from 20\% to 25\%. 
taxi driver remuneration could be assumed to be about 50% of total costs. Therefore, the total cost/km could be around 0.6 euro.

The choice of the fare level (and in parallel of the licences number) by the regulator should take into account, beside cost components, also the level of the rents that a given regulation generates on the market. An indicator of these rents is given by the fact that, despite the fact that licences have been until now assigned free of charge (with the only important exception of Bologna), they have been sold at quite high prices on the secondary market (see § 5.1).

![Figure 6: Number of taxi per 10,000 inhabitants and average cost for a urban trip of 5 km in the main Italian cities by city size.](image)

Note: Data refer to the end of 2006. Weekday urban fares in business hours. Data for the laguna of Venezia are not included.

The verification of the real returns of taxi drivers, that could be a relevant variable for the regulator, is difficult because taxi drivers are exempted from the release of a fiscal receipt. According to the data of Agenzia delle entrate, the average income declared by a taxi driver was 11,482 euro in 2004 (against 20,345 euro for a metalworker) and 13,800 in 2005. In Bologna the average income declared by the category in 2005 was 6,184 euro. These data seem to contrast with the high licences value on the secondary market. Moreover, the controls by the revenue authorities show a percentage of tax evaders in the sector of about 70%, having declared an average number of km lower of about 37% and a fuel cost lower of about 22% than the real one. Fiscal inspections have also found tax evasions in the licences sales that would amount, between 2000 and 2003, to 19.6 mln of euro in Lombardia and 18.8 mln in Lazio.24

An interesting aspect of the phase of bargaining with the municipalities following the introduction of the Bersani law is that in most cases taxi drivers asked for a fare increase as a counterpart to accept new licences. In Firenze fares have increased by more than 20% (26% and 20.8% for the fixed component and the distance charge, respectively), in Roma at the beginning of 2008 the fare increased by 18% (with a supplement of 2 euro

---

24 See *La Repubblica*, November 30, 2007, p. 3.
for departures from the Termini railway station and the suppression of the supplement for the first piece of luggage), in Milano in July 2007 there has been an increase of the distance charge of 12.6%, while the initial fixed charge has not been changed. Annual fare changes have been scheduled but subject to service quality evaluations. In Bologna in 2008 the fare increase has been about 8%; there is a discount of 3 euro (for a limited amount of trips per year) to women travelling alone between 10 p.m. and 6 a.m..

The coupling of licences increase and fare increase reflects several motivations (Aquilina, 2008). For the municipalities, it is a compromise to reduce the operators’ resistance to the increase in the number of licences. For taxi drivers, instead, the rationality of this request depends on the elasticities of demand with respect to waiting time and price, as can be derived from the model presented in § 2. From (7), if $|\omega|<1$ an increase in $n$, all other things being equal, determines a reduction in unit revenues (per operator). In this case taxi drivers could ask to the regulator a fare increase in order to try in this way to get returns back at least at the level prevailing before the increase in $n$. From (8) this happens if $|\eta|\leq 1$ (i.e. if demand is not very elastic with respect to price). In line with theoretical results, the municipality of Roma decided that fares have to be considered as “maximum fares” instead of fixed levels. This greater flexibility seems to be barely used by operators (AGSPL, 2007).

6. Some final remarks

In taxi and NCC markets demand depends not only on price but also on the quality of service. This last variable is related to general urban traffic and mobility policies and on waiting time. In a market where price and supply are fixed by a regulator, demand will adjust through changes in waiting time. This creates the possibility of multiple equilibria since demand depends on supply size through its dependence on waiting time. The regulator has a very difficult task, having to define a price and a supply level that at the same time are suitable to demand characteristics (and influence it through waiting time) and ensure an “acceptable” income to the operators in the market. This requires a continuous acquisition of information on demand and supply conditions, that was lacking in the Italian system.

Indeed the Italian regulation seems to derive more from some forms of capture of the regulator than from the presence of market failures. In local consulting bodies there is an over-representation of operators with respect to customers, with a frequent fare revision and a time-invariant supply, despite the fact that several demand indicators showed a growing dynamics. The large differences in supply, fares level and structure, and other operating rules across cities are difficult to explain on the basis of market conditions. The use of instruments of market knowledge by the regulator has been rare and sporadic, and service quality, also in terms of waiting time, has been deteriorating over time. The prohibition of creating “taxi companies” with multiple licences has no economic justification.

Instead of introducing a real deregulation, the Bersani law increased the number of options available to municipalities to increase supply, introducing also instruments of partial compensation of incumbents. The legislator’s approach seems to have acknowledged the difficulties to implement more radical reforms in a single step, given the operators’ bargaining power, especially in the big cities. The supply increase
implemented in some big cities has been obtained in exchange for fare increases that may have somehow discouraged the additional demand deriving from the reduction in waiting time.

The options chosen by municipalities among the ones available, the operators’ reactions and the effect on local regulations and markets suggest some lines for intervention that, even in a non-liberalized market, may increase efficiency in the Italian market as well in other countries with highly regulated taxi and NCC services.

First, it seems appropriate the introduction of a “higher level” of regulation in the industry. The municipal regulator hardly ever used its main advantage, the possibility of knowing closely the market, while it has often been captured by local operators (this risk being its main disadvantage). The introduction at the national level of some norms that make compulsory the use of specific indicators of supply shortage and of need of fare adjustment and the opening of the market to companies that could own more than one licence could help to improve the situation.

A second action regards the integrated government of mobility at the local level, with the use of instruments such as the increase of reserved lanes for public transport and of limited traffic zones, road pricing, integrated transport fare systems, introduction of parking fees, and so on. The improvement of these context factors could increase significantly service efficiency and average speed and reduce waiting time.

A third useful instrument could be the introduction of additional mechanisms to increase service transparency with respect both to customers and to the regulator. Among the possible measures there are a drastic reduction of supplements and the delivery to the customer of a detailed receipt (with fiscal effect too).

Last, the regulator could also evaluate the use of compensation schemes (additional to the one already set by the Bersani law) for the annulment of licences value on the secondary market, that could be used as an instrument to overcome the opposition of incumbents to further supply increases.

The Italian experience shows that regulation itself can create rents that reinforce the influence of incumbents on regulators. The resulting market setting may solve, at least in part, some of the market failures, but at a very high cost in terms of consumers’ welfare and total welfare and makes it very difficult to introduce significant changes.

Acknowledgements

I am grateful to Paolo Sestito, Magda Bianco, Silvia Giacomelli and many other colleagues of the Bank of Italy for useful comments and suggestions. The views expressed here are solely those of the author and do not necessarily reflect those of the Bank of Italy.
References


Agenzia per il controllo e la qualità dei servizi pubblici locali del Comune di Roma (2004a) *Il settore taxi a Roma. Ipotesi di riforma*.

Agenzia per il controllo e la qualità dei servizi pubblici locali del Comune di Roma (2004b) *Relazione annuale*.

Agenzia per il controllo e la qualità dei servizi pubblici locali del Comune di Roma (2004c) *Terza indagine sulla qualità dei servizi pubblici nel Comune di Roma*.

Agenzia per il controllo e la qualità dei servizi pubblici locali del Comune di Roma (2007) *Relazione annuale*.


Modeling an Integrated Public Transportation System - a case study in Dublin, Ireland

Prabhat Shrivastava 1*, Margaret O'Mahony 2

1 Professor of Transportation System Engineering
Department of Civil Engineering - Sardar Patel College of Engineering
Bhavan's Campus, Andheri (W), Mumbai – 400058, India
2 Professor of Civil Engineering, Bursar,
Head of Department of Civil, Structural & Environmental Engineering and
Director of the Centre for Transport Research, Trinity College Dublin,
Dublin 2, Ireland

Abstract

The efficiency of the public transport system in any city depends on integration of its major public transport modes. Suburban railway and public buses are the modes normally used by the majority of commuters in metropolitan cities of developed and developing countries. Integration of these two services reduces overall journey time of an individual. In this research, a model is developed for operational integration of suburban trains and public buses. The model has two sub models: a Routing Sub Model and a Scheduling Sub Model. In the Routing Sub Model, feeder routes are generated for public buses which originate from a railway station. A Heuristic Feeder Route Generation Algorithm is developed for generation of feeder routes. In the Scheduling Sub Model, optimal coordinated schedules for feeder buses are developed for the given schedules of suburban trains. As a case study the Dun Laoghaire DART (Dublin Area Rapid Transit) (heavy rail suburban service) station of Dublin in Ireland is selected. Feeder bus services are coordinated with existing schedules of the DART on the developed feeder route network. Genetic Algorithms, which are known to be a robust optimization technique for this type of problem, are used in the Scheduling Sub Model. Finally the outcome of the research is a generated feeder route network and coordinated services of feeder buses on it for the DART station.

Keywords: Coordinated schedules, Genetic algorithms, Modal integration, Optimization, Public transportation, Routing and scheduling.

1. Introduction

It has been observed that most of the metropolitan cities of developed and developing countries are facing problems due to lack of coordination among public transport facilities. Each public transport facility is planned and designed without considering its
impact on other public transport services. In fact in most of the cases these facilities compete each other instead of complementing. This unhealthy competition leads to duplication of services to many areas and hence proves to be uneconomical. Commuters have to spend more time on journeys because of higher transfer time due to lack of integration among public transport modes. The efficiency of an entire public transport system can be enhanced by overall coordination among its modes. Coordination among different modes can be achieved by system integration, which occurs at three levels: institutional, operational and physical. The literature review has revealed that many studies are carried out for optimization of services of a single mode specially bus or train but the effort is meager as far as coordination of two modes are concerned. However, routing and scheduling problems for coordinated operations were attempted by Wirasinghe (1980), Geok and Perl (1988) using analytical models. They had considered highway grid which is assumed to be rectangular and parallel to a single railway line which may not always be true in practice. They had made an attempt to describe complex transit system by approximate analytical models. Thus most of the studies on coordination of modes are limited to analytical modeling without considering a real life network (Shrivastava and Dhingra, 2000). In this research, a model is developed for operational integration of public transport modes. Development of feeder routes and schedule coordination, the two important aspects of operational integration, are attempted in this research. As a case study, Dun Laoghaire DART station is selected. Dun Laoghaire is a rapidly growing suburb of Dublin city in Ireland. The coordination between DART services and Dublin buses (public buses) at this DART station is attempted.

2. Data collection

The Dublin Area Rapid Transit (DART) is a suburban railway system in Dublin, running basically along the coastline of Dublin Bay from Greystones to Howth and Malahide. There are 32 stations on the existing DART line. Lack of coordination between public buses and DART services has been observed even during peak hours at many stations. Dun Laoghaire is one of the prominent DART stations from where large number of trips originate. It was decided to select Dun Laoghaire as the study area due to its land use pattern which allows greater scope of feeder bus services from the station. Considerable movement of commuters takes place towards many areas from the DART station.

Typical traffic surveys were conducted during the morning peak period i.e. 7 to 9 a.m. on April 28, 2004. It was observed that the maximum number of commuters travel during 8 to 9 a.m. Therefore this time period is identified as peak hour. It has been confirmed during traffic surveys that after 9 a.m. commuter traffic starts decreasing and becomes very less after 9.30 a.m. onwards. During the traffic surveys, commuters exiting the DART station were counted manually. Typical commuter counts revealed that between 8 and 9 a.m. 1293 commuters exit from the DART station. Traffic surveyors conducted sample interviews of commuters leaving the DART station. Between 8 and 9 a.m. 300 commuters were interviewed thus making a sample size above 20%. Enquiries were made regarding their destinations, mode of transport and travel time to their destinations from DART station. Commuters who did not opt public
buses for their further journeys were also asked about their willingness to shift to public buses if buses are coordinated with DART services in future. It was found that 40% of commuters have their working places very near to DART station and they have to walk even less than 5 minutes. These commuters were not interested in shifting to public buses even if they are well coordinated with DART services. The percentage of commuters willing to shift to public buses were added to those who use public buses and a potential demand matrix for public buses was developed. It was found that there are 16 destinations (nodes) for which demand exist from DART station. Table 1 indicates potential demands to various destinations. The demand for Dun Laoghaire College, Sallynoggin, Monkstown, Deans Grange, Stillorgan and Loughlinstown was found to be more than average. Thus these nodes were identified as major destinations and priority is given to these destinations for development of feeder routes. Connectivity and distances to all destinations were obtained from Dublin Street map (Dublin street map, 2000). An average speed of 15 km per hour was adopted to address the existing congestion level and road geometrics of the influence area (Scott Wilson, 2000). Using this speed, a travel time matrix was developed. The size of matrix was 17 × 17 which includes DART station and other identified 16 destinations as indicated in Table 1. The potential demand matrix and travel time matrix were used for development of feeder route network. It was also observed during traffic surveys that in the morning peak period the trains towards city centre (north bound trains) contribute about 30% passengers; the remaining 70% were by trains from city centre (south bound trains). There were nine north bound and eight south bound trains during the peak hour of 8 to 9 a.m. The schedule coordination for feeder buses is attempted for these trains during the indicated peak hour.

Table 1: Potential Demand to Various Destinations

<table>
<thead>
<tr>
<th>Node No. (code)</th>
<th>Destinations</th>
<th>Potential demand to various destinations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>7 - 8 a.m.</td>
</tr>
<tr>
<td>1</td>
<td>Dun Laoghaire DART Station</td>
<td>00</td>
</tr>
<tr>
<td>2</td>
<td>Dun Laoghaire College</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>Sallynoggin</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>Monks town</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Deans Grange</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>Temple Hill</td>
<td>02</td>
</tr>
<tr>
<td>7</td>
<td>Black Rock</td>
<td>08</td>
</tr>
<tr>
<td>8</td>
<td>Stillorgan</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>Leopards town</td>
<td>02</td>
</tr>
<tr>
<td>10</td>
<td>Foxrock</td>
<td>02</td>
</tr>
<tr>
<td>11</td>
<td>Maple Manor / Cabinteely</td>
<td>02</td>
</tr>
<tr>
<td>12</td>
<td>Lough Linstown</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>Mount Merrion</td>
<td>02</td>
</tr>
<tr>
<td>14</td>
<td>University College of Dublin</td>
<td>04</td>
</tr>
<tr>
<td>15</td>
<td>Dundrum</td>
<td>06</td>
</tr>
<tr>
<td>16</td>
<td>Sandyford</td>
<td>03</td>
</tr>
<tr>
<td>17</td>
<td>Rouches Town Avenue</td>
<td>02</td>
</tr>
</tbody>
</table>
3. Model for operational integration

The objective of the research is to develop a model for operational integration of DART services and public buses. The scope of work involves development of a feeder route network for a selected DART station, Dun Laoghaire. The feeder route network is developed in a Routing Sub Model. The next stage is schedule coordination of feeder buses for the existing schedules of DART on the developed feeder route network. The schedule coordination is attempted in the Scheduling Sub Model. Figure 1 indicates the overall methodology adopted for operational integration of two services.

Figure 1: Proposed methodologies for operational integration.
3.1 Routing Sub Model

From the literature review it is evident that heuristic approach has been very popular for development of route network. Lampkin and Saalmans (1967), Silman et al. (1974), Dubois et al. (1979), Hsu and Surti (1976), Dhingra (1980), Mandl (1980), Baaj and Mahamassani (1990 and 1995) developed bus routes using heuristic approach by insertion of nodes in base network. Heuristic approach may or may not provide optimal route structure but it is certainly able to provide good practically acceptable suboptimal solutions (Shrivastava and Dhingra, 2001). Location of various destinations (nodes), limited connectivity among some of nodes in the influence area of DART station and design of routes without further bus to bus transfer (passengers are already subjected to one transfer i.e. from DART to buses) also encouraged to use heuristic approach in this study. The heuristic algorithm described here is developed in ‘C’ language using different node selection and insertion strategies. Proposed heuristic algorithm is heavily guided by demand matrix because satisfaction of demand is one of the prime aspects for generation of routes (Baaj and Mahamassani, 1995). Thus the model for operational integration is decomposed in two sub models: one for routing and other for schedule coordination. In actual practice also user is more concerned about the waiting / transfer time rather than slightly higher journey time. It leads to higher level of discomfort and dissatisfaction if commuters have to wait longer for connecting buses to their destinations. Therefore it is decided to carry out rigorous optimization to minimize transfer time from DARTs to buses on heuristically developed feeder routes.

The proposed heuristic algorithm has two distinct parts
1. Development of shortest paths using Dijkstra’s algorithm from DART station to identified major destinations.
2. Deviation of shortest paths by inserting other identified nodes to develop feeder routes. The deviation of shortest paths has been done based on various ‘node selection and insertion strategies’.

There should be a judicious balance in satisfaction of demand due to insertion of nodes and increase in route length for development of routes (Baaj and Mahamassani, 1995). Thus the deviation of shortest paths for development of routes is governed by ‘maximum demand deviated shorter path time’ criterion. In this criterion, the deviation of shortest paths due to insertion of nodes between origin and destination is restricted to 1.5 times the travel time on the shortest paths. The nodes which are attached at the end of shortest paths are governed by ‘path extension time criterion’. The path extension time criterion fixes an upper limit on the length of routes. In the present case study this upper limit for the length of routes is kept as 15 Km (1 hr). This upper limit for the routes is decided based on the locations of various destinations identified in sample interviews which were part of traffic surveys. The upper limit on length of routes is imposed because if routes are very long then the purpose of feeder routes is lost and such routes pose difficulty in maintaining the schedules. Though the upper limit of the route length adopted in the case study is on higher side, this limit can be reduced if other DART stations are also coordinated. This is due to the fact that a particular node may have connectivity with more than one DART stations which may lead to shorter and better routes from one station as compared to other one. Traffic surveys also revealed that some of the nodes having higher demands are concentrated near the DART station and as a result many shortest paths would be developed from DART station to these major destinations. After inserting the nodes very short routes mushrooming near

32
railway station would develop. Such routes are not practically acceptable. A similar problem was felt in route generation algorithm of Baaj and Mahmassani (1995). Proposed heuristic algorithm avoids development of such routes by imposing a constraint on minimum length of shortest paths and deviation of shortest paths using node selection and insertion strategies. The various steps involved in the proposed algorithm are described as follows:

1. Prepare the demand matrix with code numbers of various nodes (destinations) to which potential demand is identified from the DART station (origin).
2. Identify the connectivity of the above nodes using the existing route map and develop travel distance matrix in kms. Non-connectivity of nodes is assigned a very high number in the matrix. Using the average speed of travel the matrix is converted into travel time matrix in terms of ‘minutes’.
3. Identify the nodes having more than average demand and select them as major destinations.
4. Develop the Shortest paths from the DART station to major destinations using Dijkstra’s algorithm.
5. Remove all the nodes from the node list which are present in any shortest path and arrange remaining nodes in the decreasing order of their demand i.e. node having highest demand is kept at top and one with least demand at the bottom. This is done so as to give priority to higher demand nodes during the insertion process. Nodes at the top are chosen first for insertion. The nodes are removed from node list because they are now the part of shortest path and hence will be the part of a route. The opportunity is given to other nodes for insertion in shortest paths / routes.
6. Identify the nodes/chain of nodes at the end of shortest paths / routes and insert them at the end of shortest path/routes using path extension time criteria. This automatically avoids delay to the higher demand nodes (major destinations).
7. The nodes, which are remaining and are already arranged as per demand, are then inserted as per node selection and insertion strategies. The lengths of routes are governed by the above mentioned time criteria which are applied depending on location of nodes and the way they are inserted in the shortest paths / routes.
8. Insertion of nodes continues one after another until all the nodes are exhausted.
9. After generating all of the routes, they are checked for backtracking. If backtracking is found and better alternatives are available they are considered and the route is suitably modified.

3.2 Node selection and insertion strategies

In development of feeder routes the nodes having higher demands should be given priority over nodes having lower demands (Shrivastava and Dhingra, 2001). Thus in node selection and insertion strategies the nodes having higher demands are given priority for insertion over lower demand nodes. The strategies adopted for insertion of any node in the shortest paths / routes are briefly mentioned below.

a) The best possible shortest path / route for any node to be inserted is first identified. The best possible shortest path / route for a particular node is decided on the basis of the ratio of saving in passengers walk time (SPWT) to increased bus passenger time (IBPT) due to insertion of the node. This ratio is calculated for all the shortest
paths / routes. The node is inserted to the shortest path / route which gives the highest value of this ratio.

Let, in fig. II
DART Station: ‘i’
Destinations having more than average demand: j₁, j₂ and j₃
Shortest paths/Routes originating from DART station ‘i’: (i,j₁), (i,j₂) and (i,j₃)
Node to be inserted: ‘k₁’
Demand from railway station ‘i’ to ‘j’: Dᵢⱼ
Demand from railway station ‘i’ to ‘k₁’: Dᵢk₁
Nodes on shortest paths/routes (i,j₁), (i,j₂) and (i,j₃) which are nearest to ‘k₁’: j₁, j₂ and j₃ (Routes/shortest paths those have no connectivity with k₁ are omitted)

Travel time on shortest path/route (i, j): tᵢⱼ
Say k₁ is inserted in route (i,j₁) the travel time will increase from ‘i’ to ‘j₁’ due to deviation of this shortest path/route. Also Dᵢk₁ passengers who had to walk for a distance of (j₁~k₁) to reach k₁ will be benefited. Therefore

Travel time from ‘i’ to ‘j₁’ via node k₁ due to its insertion: t(i,j₁) - tᵢⱼ
Increase in travel time: {t(i, j₁) - tᵢⱼ}
Delay in terms of passenger-min for bus passengers (IBPT): Dᵢj₁{t(i, j₁) - tᵢⱼ}
Walking time for passengers from j₁~ to k₁: t(j₁~k₁)
Saving in passengers-min due to walking (SWPT): Dᵢk₁t(j₁~k₁)

Calculate following (SPWT/ IBPT) ratios for all the routes as given below and consider the route for which this ratio is maximum. In this case Maximum demand deviated shorter time path criterion is adopted.

a. Dᵢk₁t(j₁~k₁) / Dᵢj₁{t(i, j₁) - tᵢⱼ}
b. Dᵢk₁t(j₂~k₁) / Dᵢj₂{t(i, j₂) - tᵢⱼ}
c. Dᵢk₁t(j₃~k₁) / Dᵢj₃{t(i, j₃) - tᵢⱼ}
b) After selecting the route / shortest path for insertion of any node the best possible way in which the node could be inserted in the selected shortest path / route is determined. The best possible way is determined on the basis of minimum additional passengers delay to successor nodes. Sometimes backtracking becomes essential at any node due to its location and connectivity with other nodes. In such cases also the above criterion of minimum additional passenger delay is used.

c) The presence of a node or series of nodes at the end of shortest path / route makes it essential to extend the route. In case of the presence of one node, both the options of inserting the node at the end of shortest path / route and between last and last but one node are analyzed. The option which gives the minimum additional passenger delay is selected. The series of nodes are attached at the end of the concerned shortest path / route if they are present at the end to avoid additional delay to higher demand nodes.

d) Sometimes due to the presence of a series of nodes near to the shortest path / route, backtracking on some nodes becomes essential. This backtracking may also increase the length of the route beyond the specified value. In such cases, to avoid backtracking and delays to higher demand nodes part of the length of shortest path is merged with the series of nodes and thus new routes are developed (Shrivastava and Dhingra, 2001).

e) Finally, all the routes are checked for undesirable backtracking. To check undesirable backtracking and to explore better options the travel time on the return journey of backtracked section is assigned a very high value. Other options, if any, are analyzed and compared with the backtracked option and the better one in terms of minimum passenger delay is selected.

3.3 Scheduling Sub Model

Attempts have been made to obtain optimal schedule on transit networks only with transfer time consideration using computer simulation (Rapp and Gehner, 1976) and combination of optimization model and simulation procedure (Bookbinder and Diesilets, 1992). However development of optimal schedules is an extremely difficult task especially for schedule coordination problem even for a small transit network. The schedule coordination problem consists of transfers between at least two modes along with other objective like vehicle operation cost or fleet size. There are constraints like keeping load factors and transfer times on various routes acceptable to both users and operators. Thus the objective function and constraints make such problems multi objective, non linear and non convex (Shrivastava et al, 2002). The difficulty due to large number of variables and constraints, the discrete nature of variables and non-linearity involved in the objective function and the constraints makes such problems difficult to be solved by traditional optimization techniques (Chakroborthy et al., 1995). In view of this, techniques like fuzzy logic have been tried for such problems (Kikuchi and Parmeswaran, 1993). Chakroborthy et al. (1995) highlighted the enormity of a similar type of problem. Even after linearizing the problem, the complexity remains very large. The benefit obtained through linearization is offset by the increase in the number of variables and constraints. In general, the number of variables and constraints required are of the order of $O(r^2n^2)$, where ‘r’ is the number of routes through a transfer station and ‘n’ is the number of buses/trains on any of the routes. Chakroborthy et al (1995) attempted to solve the linearized formulation of a similar problem, but the
algorithm failed to converge to any solution. Therefore Genetic Algorithms (GAs), which is a robust optimization technique and well suited for such problems, is applied for this phase of the research (Goldberg, 1989). The basic differences of GAs with most of the traditional methods are that GAs use coding of the variables instead of variables directly, a population of points instead of a single point, and a stochastic operators instead of deterministic operators. All these features make GAs search robust, allowing them to be applied to a wide variety of schedule coordination problems (Shrivastava and Dhingra, 2002). The following steps are involved in determination of coordinated schedules.

1. Assignment of traffic on developed feeder routes.
2. Development of objective function and constraints.
3. Calculation of penalized objective function
4. Application of Genetic Algorithm to determine optimal frequencies on different routes for minimum penalized objective function.

3.3.1 Assignment of traffic on developed routes

Potential demand to various destinations from the DART station is identified through traffic surveys. Since all the feeder routes to different destinations originate from the DART station the link connecting the station and the first node on the route is critical link. This link carries the maximum load on the route. Scheduling of buses is done on the basis of this maximum load. The assignment of traffic on feeder routes is based on the proportionate frequency criterion which is based the fact that a higher bus frequency attracts larger traffic.

3.3.2 Development of objective function and constraints

The scheduling of any public transport facility must satisfy both users and operators. The users are concerned with availability of services without waiting longer time and acceptable crowding levels. Operators are concerned with saving in operational cost of facility or minimizing the fleet size and higher crowding levels to earn profit or at least to get break even. Thus in the objective function for schedule coordination the user cost is associated with the transfer time between buses and DART services. The operator cost is taken as the vehicle operating cost which is incurred due to total distance travelled by buses (Shrivastava and Dhingra, 2002). The constraints are related to minimum and maximum load factor, minimum and maximum transfer time and unsatisfied demand. Mathematically the objective function and constraints can be presented as follows:

Objective Function:

\[
C_1 \left\{ \sum_j \sum_u \sum_l pass_j^u (bus_j^l - dart^u) \delta_j^l + \sum_j \sum_u \sum_l pass_j^u (bus_j^l - dart^u) \delta_j^l \right\} + C_2 \left\{ \sum_j f_j \cdot T_j \right\}
\]
Constraints

1. \((\text{bus}_{j}^{u} - \text{dart}^{u}) \leq T_{\text{max}}\) and \((\text{bus}_{j}^{v} - \text{dart}^{v}) \leq T_{\text{max}}\)  
   Maximum transfer time constraint

2. \((\text{bus}_{j}^{u} - \text{dart}^{u}) \geq T_{\text{min}}\) and \((\text{bus}_{j}^{v} - \text{dart}^{v}) \geq T_{\text{min}}\) 
   Minimum transfer time constraint

3. \(\frac{Q_{j}^{\text{max}}}{N_{j} \times \text{CAP}} \leq L_{\text{max}}\) 
   Maximum load factor constraint

4. \(\frac{Q_{j}^{\text{max}}}{N_{j} \times \text{CAP}} \geq L_{\text{min}}\) 
   Minimum load factor constraint

5. \(\sum_{j} d_{\text{unsat}} = 0\) 
   Unsatisfied demand constraint

Where,

\(j = \text{Number of routes available at each stations}\)

\(l = \text{Number of buses available for } u^{th} \text{ north bound DART and } v^{th} \text{ south bound DART}\)

VOC = Vehicle operating cost for Dublin buses

\(C_{1} = \text{Cost of transfer time in Euro per minute, adopted as 11.32 cents/minute for the case study, (Steer Davies, 1994)}.

\(C_{2} = \text{Cost of operation of Dublin bus per Km., adopted as € 3.66 for Dublin buses for the case study, (Scott Wilson, 2000)}.

\(\text{pass}_{u}^{j} = \text{Passengers transferring from } u^{th} \text{ north bound DART to } j^{th} \text{ route}.

\(\text{pass}_{v}^{j} = \text{Passengers transferring from } v^{th} \text{ south bound DART to } j^{th} \text{ route}.

\(\text{bus}_{j}^{l} = \text{Departure of } l^{th} \text{ bus on } j^{th} \text{ route}\)

\(\text{dart}^{u} = \text{Arrival of } u^{th} \text{ north bound DART}\)

\(\text{dart}^{v} = \text{Arrival of } v^{th} \text{ south bound DART}\)

\(\delta_{j}^{u.l} = \text{is a term which shows whether transfer of passengers is possible or not. It attains a value one if transfer from } u^{th} \text{ north bound DART to } l^{th} \text{ bus on } j^{th} \text{ route at DART station is feasible otherwise it attains a value zero.}\)

\(\delta_{j}^{v.l} = \text{is also a term which shows whether transfer of passengers is possible or not. It attains a value one if transfer from } v^{th} \text{ south bound DART to } l^{th} \text{ bus on } j^{th} \text{ route at DART station is feasible otherwise it attains a value zero.}\)

\(f_{j} = \text{Frequency of buses on } j^{th} \text{ route in terms of number of bus trips per hour}\)

\(l_{j} = \text{length of } j^{th} \text{ route in kilometers}\)

TP = Time period, hours

\(T_{\text{max}} = \text{Maximum allowable transfer time between arrival of DART and departure of connecting bus. For the case study this value is assumed as 10 minutes (Based on commuters’ opinion survey in study area).}\)

\(T_{\text{min}} = \text{Minimum allowable transfer time between arrival of DART and departure of connecting bus. For the case study this value is assumed at 5 minutes (Based on observations & opinion survey in study area).}\)

\(Q_{j}^{\text{max}} = \text{Number of passengers on first link connecting DART station on } j^{th} \text{ route for given time period.}\)

\(N_{j} = \text{Number of bus trips during entire time period under consideration (} f_{j}^{*} \text{ TP)}\)

\(\text{CAP} = \text{Seating capacity of bus, for Dublin buses it is taken as 74 (Scott Wilson, 2000)}\)

\(L_{\text{max}} = \text{Maximum load factor, it is adopted as 1.2 for the case study (Scott Wilson, 2000)}\)

\(L_{\text{min}} = \text{Minimum load factor, adopted as 1 for the case study}\)

\(d_{\text{unsat}} = \text{Unsatisfied demand}\)
The first term of the objective function involves transfer time between DART services (both nth and sth bound) and coordinating buses. The second term gives the vehicle operating cost, which is proportional to the distance traveled by buses. Constants C1 and C2 are used to convert the objective function in monetary unit of Euro (€). The first two constraints are related to transfer time (Chakraborty et al., 1995). The first constraint ensures that transfer time between arrival of a DART and departure of connecting buses should be less than a maximum value. The second constraint ensures that there should be minimum time available for transfer. This constraint is obvious because it takes a minimum time for passengers to board coordinating buses after arriving from DART. Through the traffic surveys this minimum transfer time has been established as 5 minutes. The third and fourth constraints ensure that the load factor lies within a maximum and a minimum value so that better level of service and availability of a certain minimum number of passengers can be ensured for economical operations. The maximum load factor is the ratio of crush capacity and normal capacity of Dublin buses. The crush capacity is taken as 88 and normal capacity is 74 thus the maximum load factor is taken as 1.2 (Scott Wilson, 2000). The last constraint ensures that maximum demand is satisfied and maximum number of commuters get coordinating buses during the period of analysis (Shrivastava et al, 2002). None of the above constraints are rigid. These constraints are obeyed and violated as per their relative importance and magnitude is directly proportional to potential demand associated with a particular constraint. Penalties are decided as per the extent of violation of constraints i.e. higher penalties are imposed for greater violation of these constraints.

### 3.3.3 Calculation of penalized objective function

The objective function and constraints as mentioned above pose a constrained optimization problem. Transformation methods are the simplest and most popular optimization methods of handling constraints. The constrained problem is transformed into a sequence of unconstrained problems by adding penalty terms for each constraint violation. If a constraint is violated at any point, the objective function is penalized by an amount depending on the extent of constraint violation (Deb, 1995). Three sets of penalties are decided which are added to objective function and penalized objective function is calculated. The following penalties are used in analysis:

1. Transfer time penalty
2. Load factor penalty
3. Penalty for unsatisfied demand

These penalties are function of objective function, penalty coefficient, number of affected commuters and adopted bus capacity.

#### 3.3.3.1 Transfer time (tt) penalties

As stated above, it is observed during the surveys that it takes about 5 minutes on average to reach a bus stop after arriving from the DART. Thus the minimum transfer time from DART to bus is adopted as 5 minutes. Therefore, any bus which starts after 5 minutes of the scheduled arrival of DART is considered as a connecting bus to that particular DART service. A transfer time between 5 to 10 minutes is regarded as acceptable. In fact considering 5 minutes as the minimum time required for transfer, effective waiting time lies between zero to five minutes which is considered as acceptable. Any transfer after 10 minutes i.e. effective waiting time more than 5 minutes
is penalized. Higher values of penalty coefficients are adopted for higher transfer time because higher transfer time causes more discomfort to passengers.

3.3.3.2 Penalty due to unsatisfied demand

If some passengers are not able to get any bus in the specified duration of analysis then it is taken as unsatisfied demand and the penalty is imposed on objective function.

3.3.3.3 Load factor (LF) penalties

The minimum value of load factor is adopted as ‘1’. The value of maximum load factor is adopted as 1.2 so as to maintain a better level of service. Level of service becomes poor due to a rise in load factor above maximum adopted value. If the load factor becomes less than ‘1’ it leads to uneconomical operation which may not be acceptable to operators. Therefore higher values of penalty coefficients are adopted as load factor increases above the maximum specified value similarly higher values are adopted as load factor decreases below minimum value.

3.3.4 Application of Genetic Algorithms

In the real world, the process of natural selection controls evolution. Organisms most suited for their environment tend to live long enough to reproduce, whereas less suited organisms often die before producing young or produce fewer and/or weaker young. In the applications of Genetic Algorithms process of evolution is studied by creating an artificial world, populating it with pseudo organisms and giving those organisms a goal to achieve (Goldberg, 1989). Genetic Algorithms store the characteristics of artificial organisms in a Genotype, which mimics the DNA of natural life. The genotype is nothing more than a long string of bits. A bit is the smallest piece of data a computer can process. It can be only one of two values: ‘0’ or ‘1’. A bit in the genotype string can be ‘on’ which has the value ‘1’, or can be ‘off’ which has the value ‘0’. The existence of a certain characteristic can be indicated by whether a particular bit is set to ‘on’ or ‘off’. The operation of GAs begins with population of random strings representing design of decision variables. Thereafter, each string is evaluated to find the fitness value. The population is then operated by three main operators- reproduction, crossover and mutation to create a new population of points. The new population is further evaluated and tested for termination. If the termination criterion is not met, the population is iteratively operated by the above three operators and evaluated. This procedure is continued until the termination criterion is met. One cycle of these operations and subsequent evaluation procedure is known as a ‘generation’. The GAs use search strategies by using probability in all their operators. Since an initial random population is used, to start with, the search can proceed in any direction and no major decisions are made in the beginning. Later on, when the population begins to converge in some bit positions, the search direction narrows and optimal or near optimal solution is achieved. Thus nature of narrowing the search space as the search progresses is adaptive and is unique characteristic of Genetic Algorithms (Deb, 1995). Therefore Genetic Algorithms always guarantee the optimum / near to global optimum solution for ill behaved functions. Solutions even near to global optimum obtained by GAs are acceptable for practical problems, like the one which is being attempted in this research.

‘Reproduction operator’ is usually the first operator applied on a population. Reproduction selects a good string in a population and forms a mating pool. In the
'Crossover operation', information among strings of the mating pool is exchanged and new strings are created. 'Mutation' adds new information in a random way to the genetic search process, and ultimately helps to avoid GAs from getting stuck at local optimums. In the present analysis 'uniform random' and 'roulette' selection operators are compared. Similarly ‘simple’ and ‘uniform’ crossover are tested. Best among ‘simple invert’, ‘simple random’ and ‘swap’ mutation is used (Lance Chambers, 1995).

3.4 Use of Genetic Algorithms for objective function and constraints

The above objective function is used with LibGA software (Lance Chambers, 1995) of Genetic Algorithms in Linux environment to determine optimal frequencies on developed feeder route network. Genetic Algorithms parameters are tuned for the objective function and thus type of process and best values of operators are decided. The following are the outcomes of several runs for tuning Genetic Algorithms parameters.

- Roulette and uniform random selections are compared and it is found that Roulette selection converges faster for our objective function.
- Simple random and Swap mutation give better results as compared to Simple invert. In the analysis the Simple random mutation is adopted.
- Uniform crossover converges earlier to Simple crossover. Thus uniform crossover is adopted for the analysis.
- Among seed values 1 to 10 seed value ‘1’ gave best results and hence is adopted for analysis.
- The value of penalized objective function for pool size 30 is found to be same as obtained for pool size 70 and above. Therefore pool size 30 is adopted which has the advantage of lesser computational time also.
- It is found that combination of crossover probability of 0.85 and mutation probability of 0.005 gave the lowest value of penalized objective function. Thus these values are used for the analysis.

Using the above Genetic Algorithm parameters, a set of penalty coefficients for transfer time, load factor and unsatisfied demand are decided. The coefficients are decided so as to keep the load factor in the range between 1 (minimum load factor) and 1.2 (maximum load factor), the percentage unsatisfied demand as low as possible and the effective waiting time for larger percentage demand between ‘zero’ and ‘five’ minutes. The demand satisfaction and load factors on various routes are two dominating factors for both users and operators. It has been found during the interviews of commuters that they prefer to have connecting buses with in five minutes of waiting after arriving at bus stops but most of them even accept ten minutes of waiting as a reasonable time. Thus the variation of penalty coefficients for minimum load factor is studied on percentage satisfaction of demand with in ten minutes of waiting. The coefficient for minimum load factor is selected because it is observed that the load factor frequently goes below 0.4 (minimum value) due to low demand which is not compatible to adopted existing bus capacity. Table 2 indicates typical variation of overall load factor (average load factor of all the routes), percentage demand satisfied with in ten minutes of waiting and values of penalized objective function. This typical variation is observed when penalty coefficient corresponding to minimum load factor (less than 0.4) is varied keeping other coefficients same. The typical variation in the table shows that Genetic Algorithms are very sensitive to penalties. A weighted factor is
calculated by awarding equal weights to the overall load factor and percentage demand satisfaction with in ten minutes of waiting. Penalty coefficient corresponding to higher weighted factor is selected for further analysis.

Table 2: Typical variation of over all Load factor, satisfied demand with in ‘10’ minutes of waiting and penalized objective function with respect to coefficient of minimum load factor penalty

<table>
<thead>
<tr>
<th>Value of Coefficient for minimum load factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over all load factor (average for all the routes)</td>
<td>0.3074</td>
<td>0.3074</td>
<td>0.3656</td>
<td>0.3656</td>
<td>0.4210</td>
<td>0.5055</td>
<td>0.5055</td>
<td>0.5141</td>
<td>0.5310</td>
<td>0.5423</td>
</tr>
<tr>
<td>% demand satisfied with in 10 minutes of weighting</td>
<td>95.24</td>
<td>95.24</td>
<td>90.04</td>
<td>86.77</td>
<td>85.11</td>
<td>77.11</td>
<td>76.19</td>
<td>75.19</td>
<td>73.34</td>
<td>71.74</td>
</tr>
<tr>
<td>Typical values of penalized objective function</td>
<td>61804</td>
<td>81708</td>
<td>92732</td>
<td>105126</td>
<td>125158</td>
<td>132600</td>
<td>140042</td>
<td>149857</td>
<td>155529</td>
<td>173348</td>
</tr>
</tbody>
</table>

The penalties discussed above are calculated using the selected set of penalty coefficients and the penalized objective function is determined by adding penalties to the objective function. A set of frequencies on various routes corresponding to the minimum value of the penalized objective function is used for determination of coordinated schedules on various routes.

4. Results and discussion

It was found that there are 6 destinations having demand greater than average. These destinations are Dun Laoghaire College, Sallynnoggin, Monkstown, Deans Grange, Stillorgan, and Loughlinstown. Using Dijkstra’s algorithm, four shortest paths were developed. These shortest paths were modified by node selection and insertion strategies and four feeder routes were obtained. The developed feeder route network is shown in Figure 3 with the code numbers of nodes as given in Table 1. The lengths of feeder routes 1, 2, 3 and 4 are 5.54, 9.10, 14.6 and 5.8 km respectively. If similar exercise is carried out by identifying influence area of all stations shorter feeder routes will be developed. This is due to the fact that one node may be connected to more than one DART station and its connectivity will certainly be better with shorter connecting length from one particular station. This will lead to smaller feeder routes which will ultimately help in maintaining schedules of feeder buses (Shrivastava and Dhingra, 2001). It can be seen in Figure 3 that destinations like Stillorgan (8), Mount Merrion (13), University College Dublin (14) and Dundrum (15) are closer to Blackrock DART station as compared to Dun Laoghaire. Thus feeder routes for these destinations from Blackrock will be shorter. In the existing route structure of Dun Laoghaire bus routes numbers 46A, 75, 111, 59, 46X originate from station where as route numbers 7, 7A and 45A pass through the station with origins elsewhere. Some of the existing routes that originate at the station pass through some of the locations for which demand does not originate from the station as indicated in our typical traffic survey. The route
Number 75 passes through Stillorgan, Leopardstown, Sandyford, Ballinteer, Oldbawn and Tallaght.

Our typical survey shows the demand for last three destinations is nil from the station. Moreover the length of this route is very long having existing trip time more than one hour. Such longer routes pose problems in maintaining schedules. Similarly route number 46A goes to the city centre thus duplicating the services of the DART towards city centre. Route number 45A goes to Bray which is parallel to the DART line. Route number 59 passes through Dalkey and Killiney. Route number 111 also passes through Dalkey and goes to Loughlinstown. The routes 59 and 111 could have been clubbed together and a single feeder route could have served the purpose. Thus it can be concluded that the existing routes do not serve the purpose of feeder routes and lead to duplication of services. Table 3 gives typical details of bus schedules with load factors on different routes during morning peak hour. The average load factors on 2\textsuperscript{nd}, 3\textsuperscript{rd} and 4\textsuperscript{th} routes are more than 0.4 and the overall load factor for all the routes is 0.3650. Average load factor on route ‘1’ is very low and this is due to the fact that during the hour of analysis the potential commuters for destinations lying on this route are less as compared to adopted existing capacity of buses in the analysis. It will be appropriate to use buses with smaller capacity on such routes. Moreover the local demand which is not considered at various nodes will further improve the load factors. The local demand is not considered because the routes are designed for feeder buses from DART station. Hence satisfaction of demands which generate from DART station is of prime concern. The load factors can be further improved if DART schedules are optimized beforehand.
(Shrivastava and Reddy, 2002). However if DART schedules are modified then coordination for other direction travel i.e. from buses to DART should also be studied. In the existing scenario due to frequent availability of DART services there will always be coordination from buses to DARTs irrespective of arrival time of buses.

Table 3: Details of Bus Schedules with Load Factors

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Train Timings</th>
<th>Bus Timings</th>
<th>Load Factors</th>
<th>Over all load factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>North Bound DARTS</td>
<td>South Bound DARTS</td>
<td>Route 1</td>
<td>Route 2</td>
</tr>
<tr>
<td>1</td>
<td>08.08</td>
<td>08.02</td>
<td>8.07</td>
<td>8.07</td>
</tr>
<tr>
<td>2</td>
<td>08.15</td>
<td>08.09</td>
<td>8.22</td>
<td>8.13</td>
</tr>
<tr>
<td>3</td>
<td>08.23</td>
<td>08.20</td>
<td>8.37</td>
<td>8.19</td>
</tr>
<tr>
<td>4</td>
<td>08.29</td>
<td>08.25</td>
<td>8.52</td>
<td>8.25</td>
</tr>
<tr>
<td>5</td>
<td>08.33</td>
<td>08.31</td>
<td>8.31</td>
<td>8.55</td>
</tr>
<tr>
<td>6</td>
<td>08.38</td>
<td>08.36</td>
<td>8.37</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>08.43</td>
<td>08.45</td>
<td>8.43</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>08.49</td>
<td>08.53</td>
<td>8.49</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>08.58</td>
<td>-</td>
<td>8.55</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>Trains after 9 a.m.</td>
<td>Buses to be scheduled after 9 a.m.</td>
<td>Buses to be scheduled after 9 a.m.</td>
<td>Load factor for Buses to be scheduled after 9 a.m.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4 gives waiting time details corresponding to developed coordinated schedules. It can be seen from the table that 65.07% demand is satisfied within ‘0’ to ‘5’ minutes of waiting and 24.97 % of demand is satisfied between ‘6’ to ‘10’ minutes of waiting. Thus a total demand of 90.04% is satisfied within ‘10’ minutes of waiting. Entire demand is satisfied before ‘15’ minutes of waiting. In the present scenario since the existing routes do not serve the purpose of feeder routes average waiting time of commuters at Dun Laoghaire DART station is more than 15 minutes even during the morning peak hour with load factors in the range of 0.2 to 0.3.

Table 4: Waiting Time Details of Passengers

<table>
<thead>
<tr>
<th>Duration of Delay in Minutes</th>
<th>Route No : 1</th>
<th>Route No : 2</th>
<th>Route No. 3</th>
<th>Route No.4</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 5</td>
<td>42.8</td>
<td>87.3</td>
<td>65.2</td>
<td>65.0</td>
<td>65.07</td>
</tr>
<tr>
<td>6 - 10</td>
<td>32.2</td>
<td>12.7</td>
<td>27.5</td>
<td>27.5</td>
<td>24.97</td>
</tr>
<tr>
<td>11 - 15</td>
<td>25.0</td>
<td>nil</td>
<td>07.2</td>
<td>07.5</td>
<td>09.96</td>
</tr>
<tr>
<td>More than 15</td>
<td>nil</td>
<td>nil</td>
<td>nil</td>
<td>nil</td>
<td>nil</td>
</tr>
</tbody>
</table>
5. Conclusions

Following conclusions can be drawn from this research.

− In this research, the model has been developed for operational integration of two services i.e. public buses and a rail service (DART) for one DART station only. The same modeling exercise can be repeated at various other DART stations after identifying influence area of each for different time periods of a day. Thus the model can develop an integrated public transport system in which suburban trains / DART services will work as main line haul service and buses can feed the local areas. This type of integrated system will allow both the modes to compliment each other instead of competing. The integrated system will also reduce wasteful duplication of services. If the modelling exercise is repeated to other DART stations route structures will be better in terms of lengths and satisfaction of demands.

− It is also confirmed that Genetic Algorithms are very efficient in solving multi objective, non linear schedule coordination problem. The time taken to obtain results is directly proportional to adopted population size. Near optimal results can be obtained with smaller population sizes, which take less computational time and are practically acceptable in real life situations. In the case study population size 30 is selected which takes less computation time and is able to provide equally good results as provided by higher population sizes.

− The variation of percentage demand satisfaction and overall load factor against variation of minimum load factor penalty shows that the Genetic Algorithms are very sensitive to penalties. Thus selection of appropriate penalties is very much required before the optimization process.

− The model developed in the research considers and develops real life network with real life objectives for both users and operators. The model takes into account real life constraints like level of service (maximum load factor), economical operation (minimum load factor), minimum and maximum transfer time and availability of public buses to maximum number of commuters (constraint for unsatisfied demand). The model maintains a judicious balance between load factor and satisfaction of demand within acceptable waiting time. Thus the model is able to provide satisfactory results (feeder routes and coordinated schedules) from users and operators point of view. Hence it can be claimed that proposed modeling exercise is a specific contribution towards realistic modeling on coordinated operations for passenger trips.

Acknowledgements

The research is funded under the Programme for Research in Third-Level Institutions (PRTLI), administered by the HEA. The authors would like to acknowledge the useful discussions with Dr Biswajit Basu during the research work.
References


Abstract

This paper addresses the impact of information and communication technology (ICT) on sustainable transport by examining the direct application of ICT in urban transport. Following a discussion of various negative externalities of transport, the paper examines the extent to which existing and potential ICT applications in the transport sector can assist in making urban transport more sustainable than it is at present. The focus of analysis is on qualitative and quantitative impacts of several ICT applications on travel behaviour (including fatalities), factors that influence adoption, and impacts of adoption including potentially secondary effects. The literature suggests that ICT innovations are most effective in fatality reduction, but it seems that these are also quite effective in reducing fuel consumption through fuel-intelligent vehicles.

Keywords: Information and communication technology; Urban transport; Sustainability; Excessive driving; Congestion relief; Fatality reduction; Fuel-intelligence.

1. ICT applications and the road to sustainable transport

Car traffic in urban areas uses 50% more energy than car traffic in non-urban areas. Therefore, it makes sense to specifically focus on urban transport in the context of the need for an increased sustainability of transport. Various technology options are open to bring sustainability aims in transport nearer, like the use of electric vehicles, hybrid vehicles deriving electricity from hydrogen operated fuel cells or batteries, and the use of vehicles that are cleaner and more efficient in using fossil fuels. Many studies are devoted to the energy road towards sustainable transport, but only a few are concerned with the use of novel ICT applications.

This paper addresses the role of information and communication technology (ICT) in supporting urban transport to become more sustainable. It examines potential

* Corresponding author: Marina van Geenhuizen (m.s.vangeenhuizen@tudelft.nl)
sustainability increases but also critical hurdles that prevent the adoption of particular applications and achievement of the right behavioural responses. We focus on qualitative and quantitative impacts of several ICT applications on travel behaviour, including fatalities. Broadly speaking, information and communication technology is a set of heterogeneous technologies (hardware and software) that allow for electronic communication, data collection and processing in distributed networks, and electronic guidance and management through sensor technologies. ICT applications in the transport system differ in complexity, ranging from simple electronic communication (signals) to interactive and highly intelligent applications in traffic management and control, and in car fuel management. Differences in complexity become also apparent if one considers the position of the new technology in the urban transport system perceived as a system of layers, i.e., infrastructures, services on these infrastructures, vehicles moving through the system of infrastructures, and persons and freight moving in these vehicles. Of course, from a holistic perspective, it would be necessary to also consider the extent in which the electronic devises used in the different solutions are manufactured in a sustainable way, like electric road screens and in-vehicle electronic devices and sensors. However, this perspective falls beyond the current study.

Sustainability of the transport sector in urban areas is a major concern today for governments throughout the developed economies of the world. Although the sustainability concern was focused on negative environmental externalities of the transport sector in the early 1990s, the term has a much broader meaning today. Concerns over greenhouse gas emissions and global climate change, as well as the potential depletion of petroleum, the world’s major transport fuel, have been joined by concerns for urban air quality, excessively large numbers of vehicle accidents and their resulting fatalities and injuries, and congestion.

All of the major externalities related to non-sustainability are a function of traffic volume. Thus, in order for the transport system to become sustainable all that is necessary is to decrease the amount of transport consumed. The paper looks to ICT innovations in the transport system as possible ways of accomplishing the outcome of sustainability. To this purpose, we first introduce the layer model of urban transport systems. A main difference between ICT applications is their position in the system in that some of them are fixed (or semi-fixed) in the physical infrastructure (first layer), e.g. automated guideways (physical), on-road signing including variable messaging, and surveillance systems, and others are mobile in the sense that they are in-vehicle systems or personal (portable) systems. Another main difference between ICT applications resides in the type and number of layers in the urban transport system required to implement the application. Thus, some applications make use of one layer, while others make use of two or three of them. If more layers are involved, the technical complexity seems higher because of the additional infrastructure requirements; it also implicates the involvement of a larger number of different transport actors, potentially increasing social complexity and a delay in the adoption of the innovations concerned. ICT innovations in the transport system may also be categorized according to the role of the information concerned, ranging from information to support choices of car drivers and passengers, to information that serves to take over drivers’ decisions. The last role is causing legal issues concerning responsibility and liability to enter the scene. Such a situation tends to delay the adoption of the innovations concerned.

In the paper, a closer look is taken at various practical ICT innovations that can be used to reduce non-sustainable fuel use and other negative externalities by
distinguishing between three sustainability areas, that is: (1) excessive driving, (2) congestion relief and (3) fatality reduction. We will focus on private cars (persons and freight) and public transport using roads in urban areas. For each of the applications addressed we discuss to what extent the applications are currently adopted and what the effectiveness is in terms of impact on sustainable transport. In the last context, attention is also given to compensation and circumvention behaviour of drivers causing a reduction of the overall impacts in the urban transport system. The source of analysis is a compilation of the recent international literature on transport behaviour and transport policy (last three years).

The paper is structured as follows. After a brief discussion of transport non-sustainability, a closer look is taken at the urban transport system concerned by examining ways in which applications of ICT in various layers of this system can lower (or alter) negative impacts of using the transport system. A distinction is made between excessive driving, congestion relief and fatality reduction. Finally, differences in the potential for ICT applications are indicated and positive impacts on sustainable urban transport are evaluated by considering appropriate behavioural response and by considering higher levels in the urban transport system than the level directly affected by the applications.

2. Transport non-sustainability and the application of ICT

Greenhouse gas emissions that lead to global warming, emissions leading to air pollution with its negative impacts on human health, the large number of fatalities and injuries, diminishing petroleum reserves, and congestion, are generally accepted as components of non-sustainability by most scholars working in this area today. They are the factors that will prevent future generations from carrying out transport in the same manner that the current population does, and in effect, this is what leads to non-sustainability (e.g. Black, 2003). Before examining the various ICT impacts on problems of sustainable transport it is appropriate to state what may be a very obvious relationship. It is that all of the major externalities related to non-sustainability are a function of or are influenced by the volume of traffic (Black and Van Geenhuizen, 2006). Stated quite simply the higher the vehicle miles or kilometers traveled in a country, the greater the level of total greenhouse gas emissions from the transport sector of that country. This is true of the criteria pollutants such as carbon monoxide, sulphur oxides, nitrogen oxides, and others, as well as traditional greenhouse gases, such as carbon dioxide. Motor vehicle accidents are also a function of traffic volume. This is so obviously the case that when researchers begin to examine these, one of the first things they do is divide accidents by flow volume and work with the resulting rate of accidents per million miles traveled or some similar measure. The use of gasoline, and therefore the eventual depletion of petroleum stocks, also increases with flow volume, not the volume on a single road, but the total vehicle miles travelled. It is certainly true that different vehicles and vehicle models differ in their fuel economy, but in general the greater the vehicle miles driven by a country’s motor vehicle fleet, the greater the amount of fuel that is used. Finally, as flow volumes increase, they begin to exert influences on other vehicles in the traffic stream and this creates congestion. The ECMT has defined congestion as “the impedance vehicles impose on each other, due to speed
flow relationships, in conditions where the use of the transport system approaches capacity.” (ECMT, 2000, p. 220).

A major way to achieve a sustainable transport system is to decrease the amount of transport consumed. This is much easier to state than it is to accomplish. Regulatory policies could be formulated to achieve this end, or nations could encourage voluntary actions to accomplish the same end. It goes without saying that the former is very unpopular and the latter is unsuccessful. Therefore, various ICT innovations in the transport system are examined as possible ways of accomplishing the same outcome. To this end, first a simplified layer model of the urban transport system is introduced that applies to persons and freight transport, and clarifies the different actors involved and the targeted aspects of the system (Figure 1).

<table>
<thead>
<tr>
<th>Layers</th>
<th>Actors involved</th>
<th>Target of ICT innovation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer 4. Persons and freight</td>
<td>Private car users</td>
<td>Drivers’ behaviour: route selection, driving speed, reaction in driving, reduction of driving tasks.</td>
</tr>
<tr>
<td>Persons (drivers), parcels, containers, bulk, etc.</td>
<td>Public transport users</td>
<td>Passenger behaviour in public transport: mode choice and route selection.</td>
</tr>
<tr>
<td></td>
<td>Freight transport users</td>
<td>Quick first aid after accident.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Freight: routing and load matching.</td>
</tr>
<tr>
<td>Layer 3. Vehicles moving through the system</td>
<td>Owners of private vehicles</td>
<td>Size of flow, speed of flow, identification of obstacles, in-between vehicle distance in flow (longitudinal, lateral), collision avoiding.</td>
</tr>
<tr>
<td>Trains, cars, busses, vans, bikes, vessels, etc.</td>
<td>Logistics providers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chain organizers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vehicle manufacturers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICT manufacturers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Public transport companies</td>
<td></td>
</tr>
<tr>
<td>Layer 2. Services on the infrastructure</td>
<td>Public transport companies</td>
<td>Providing/preventing (or slowing down) access of public transport services to persons</td>
</tr>
<tr>
<td>Public transport services (time schedules) services for maintenance and transport management</td>
<td>Operators of links and nodes</td>
<td>Improvement of matching different services (seamless connections)</td>
</tr>
<tr>
<td></td>
<td>ICT system manufacturers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Public transport companies</td>
<td></td>
</tr>
<tr>
<td>Layer 1. Physical infrastructure (links, nodes)</td>
<td>Infrastructure providers</td>
<td>Providing/preventing access to infrastructure links and nodes to vehicles</td>
</tr>
<tr>
<td>Rail, road, airline, pipelines, waterways, etc.</td>
<td>Infrastructure owners</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Public authorities</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: A simplified layer model of the urban transport system.

The model comprises four layers, i.e., physical infrastructure, services to let the infrastructure work, vehicles moving through the infrastructure system, and persons and freight using these vehicles. In fact, a fifth layer can also be distinguished, that is the layer of institutions that influence regulations, arrangements between actors involved and legal and liability issues in case of accident and fatalities, etc. Impacts of interest for our analysis are mainly generated in the layer of vehicle flow (3) and the layer including persons (drivers) and freight (4). The main difference between ICT applications is that some of them are fixed (or semi-fixed) in layer 1 as they form part of the physical infrastructure, e.g. automated guideways (physical), on-road signing including variable messaging, surveillance systems, and on-road access and charging systems, while others
are *mobile* in the sense that they are in-vehicle systems or personal (portable) systems. Another main difference between ICT applications resides in the type and number of layers required to implement the application. Thus, some applications make use of one layer, while others make use of two or three of them. If more layers are involved, particularly including layer 1, the technical complexity is greater, because of the additional infrastructure requirements, and investment levels are higher with higher user costs. A situation of more layers also implies the involvement of a larger number of different actors with different interests, and different problem-definitions and –solving; a situation potentially leading to delay in the adoption of the innovations concerned.

ICT innovations in the transport system can also be categorized according to the role of the information concerned in drivers’ behaviour. A distinction can be made between information that:

a. *Supports choices* of car drivers and passengers, e.g. on-road information on upcoming congestion, routing advice from a navigation system to avoid congestion, driving advice to optimize the use of car fuel, or the real arrival time of public transport busses.

b. *Reduces options or limit drivers’ behaviour*, e.g. avoiding parts of networks, or limits to driving speed.

c. *Alerts drivers or passengers* without constraining behaviour, e.g. various modes of advanced driver assistance, like collision avoidance and lane keeping systems.

d. *Serves to take over drivers’ decisions*, fully or partly, like in electronic bonding of cars and in intelligent speed adaptation and intelligent fuel use adaptation.

The above roles played by information show different degrees of constraints upon drivers’ free choices. A special case are ICT innovations that takeover drivers’ decisions, because legal issues concerning responsibility and liability enter the scene and these still need to be settled, like responsibility and legal liability of drivers, ICT system manufacturers and the operators of network systems. This situation acts like a barrier and tends to delay the adoption of the innovations concerned.

3. A closer look at ICT innovations

Below several types of ICT innovations are examined that can be used to address the negative externalities noted above by distinguishing between the three sustainability areas, excessive driving, congestion relief and fatality reduction (Table 1). The focus of analysis is on private cars (persons) and public transport using roads in urban areas. For each of the applications addressed will be indicated whether it is fixed or in-vehicle (mobile), and whether it concerns private or public transport or both. Note that some applications show overlap because the systems may serve for example both reduction of excessive driving (reduction of fuel consumption) and congestion relief.

3.1 Excessive driving

Excessive driving contributes to fuel utilization, as well as the generation of air toxics detrimental to urban and global environments. The technologies intended to decrease the need for travel or to increase the efficiency of travel that does take place, to be
discussed are signalization, navigation systems and ICT-based intelligent vehicles with reduced fuel consumption.

**Signalization**
ICT can decrease fuel use by increasing the efficiency of the movement that does take place. This outcome can be accomplished through improved signalization. Such signalization can be phased in some areas and demand responsive in other areas, but the objective is to decrease the amount of vehicle standing time while the motor is running. Under the former the traffic signals are set so that signals in a series will change at a set frequency so that the vehicle does not have to stop. In the latter case the signals will change in response to a vehicle approaching a sensor in the roadway. Signalization – both phased and demand responsive – are widely applied. In a more advanced mode, signalization forms part of Vehicle Guiding Systems aimed at the creation of continuous flow at certain sections of roads without stops.

**Navigation Systems**
Geographic positioning systems (GPS) in conjunction with geographic information systems (GIS) offer the possibility of decreasing the amount of time spent on search behavior by motorists. Assuming one inserts his/her origin and destination to the system, the shortest route will be proposed. Such navigation systems are already common today, either portable or fixed (in-built in the car). In an alternative mode, an increasing number of motor vehicles will undertake the way finding for you and minimize unnecessary travel. The use of mobile communication in route advising seems underestimated for private car use and deserves more attention (Townsend, 2004). It is obvious that for privacy reasons, this kind of systems is not yet popular among private car drivers (see, e.g., Lee-Gosselin, 2002).

Systems that optimize route choice have seldom the primary aim of reducing the environmental effects of driving (lowest total fuel consumption) instead of the traditional aim of shortest time or distance. In a study of real traffic driving patterns in the city of Lund (Sweden), the most fuel-economic route was extracted and compared with the original route choice (Ericsson et al., 2006). It was found that the drivers’ route choice produced trips that could save 8.2% fuel by using a fuel-optimized navigation system. This corresponded with 4% fuel reduction for all journeys longer than 5 min. in Lund. Whether a fuel-optimized routing option can be included in existing navigation systems, how drivers in reality respond to the fuel-saving outcomes and what the fuel reduction turns out to be, remains unknown and will be clarified in future research steps.
Table 1: ICT applications, aims and effectiveness.

<table>
<thead>
<tr>
<th>Application</th>
<th>Aim</th>
<th>Adoption</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excessive driving reduction</td>
<td>-Decrease of vehicle standing time with running motor</td>
<td>Broad</td>
<td>No available data, but seems effective.</td>
</tr>
<tr>
<td>-Signalization (e.g. set in series) (fixed), (private, public)</td>
<td>-Decrease of search time while driving</td>
<td>Quickly, increasing</td>
<td>No available data; seems effective in time but maybe longer journeys.</td>
</tr>
<tr>
<td>-Navigation systems (in-vehicle) (private, public)</td>
<td>-Fuel-efficient routing</td>
<td>No/limited</td>
<td>8.2% fuel saving (Sweden)</td>
</tr>
<tr>
<td>-Fuel-intelligent vehicles (in-vehicle) (private, public)</td>
<td>-Prevent stop-start behaviour</td>
<td>No/limited</td>
<td>33% fuel economy improvement.</td>
</tr>
<tr>
<td>Congestion relief</td>
<td>-Monitors obstacles in the road network and sends help</td>
<td>Broad, on critical links and nodes</td>
<td>No available data, but seems effective.</td>
</tr>
<tr>
<td>-Video Surveillance and Response (fixed) (public, private)</td>
<td>-Give information on changing road network conditions ahead (persons and freight)</td>
<td>Increasingly, on critical sites</td>
<td>Overall travel time reduction by 1-2% in regular congested areas (EU).</td>
</tr>
<tr>
<td>-Variable Message Signs (VMS) (fixed) (public, private)</td>
<td>-Give customized information Limited on network conditions ahead.</td>
<td>No available data.</td>
<td></td>
</tr>
<tr>
<td>-Advanced Traveller Information Systems (mobile) (private).</td>
<td>-Supports the longitudinal following task to reduce variation in acceleration and waiting time.</td>
<td>Limited but increasing</td>
<td>Reduces variation in acceleration by 40-50% (EU). ISA (a) reduces fuel use by 8% (UK).</td>
</tr>
<tr>
<td>-Advanced Drivers’ Assistance (ADAS) (cruise control, speed adaptation) (in-vehicle) (private).</td>
<td>-Adapts speed. Information to reduce waiting time and searching time (mobile communication).</td>
<td>Limited to increasing</td>
<td>No available data, but seems effective.</td>
</tr>
<tr>
<td>-Dedicated Short Range Communication (between following and oncoming cars) (mobile) (public, private).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatality reduction</td>
<td>-Reduce waiting time for assistance after accident.</td>
<td>Limited (up-market)</td>
<td>No data available, but seems effective.</td>
</tr>
<tr>
<td>-Accident Sensors (in-vehicle) (private).</td>
<td>-Alert drivers of cars behind, alert drivers on blind spot and obstacles during night.</td>
<td>Limited (up-market)</td>
<td>No data available, but seems effective.</td>
</tr>
<tr>
<td>-Extended Viewing Systems (radar, sensors, infra-red) (in-vehicle) (private).</td>
<td>-Gives advice on (or enforces) speed reduction.</td>
<td>Broad</td>
<td>Decrease of speed, but compensation behaviour (US); reduction of crash potential by 5-17% (Canada).</td>
</tr>
<tr>
<td>-Speed Advisory/Control (fixed/variable) (private, public).</td>
<td>-Controls positioning of vehicle (lane, vehicles, obstacles) and adapts speed.</td>
<td>Limited (up-market)</td>
<td>Reduction of fatalities and heavy injury up to 30-38% (dependent on road type) (NL).</td>
</tr>
<tr>
<td>-Advanced Drivers’ Assistance (ADAS) (in-vehicle) (private).</td>
<td>-Fixes vehicle position in a flow at constant speed and distance (physical and electronic systems).</td>
<td>No</td>
<td>Several gains expected.</td>
</tr>
<tr>
<td>-Automated guided vehicles (fixed/in-vehicle) (private, public).</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: (a) ISA: Intelligent Speed Adaptation.
Fuel-Intelligent vehicles

The demand for fuel-efficient cars has been growing in the previous years driven by the increased and still increasing price of oil. Hybrid electric cars have already found presence in the marketplace due to the promise of saving fuel by using an electric motor in place of the internal combustion engine during particular stages of driving. All the major car manufacturers have developed or are currently developing hybrid vehicles, with the earlier models being relatively small, like the Toyota Prius and Honda Insight and Civic, and larger model being currently released, like a hybrid Ford Escape and various Lexus models. Despite fuel savings, the primary disadvantage of the hybrid technology from an adoption perspective is the initial cost for consumers that can be as much as 70% more than an equivalently powered internal combustion engine-only vehicle (Manzie et al., 2007). At the same time, fuel-intelligent cars are being developed equipped with a relatively cheap sensor network.

Fuel consumption in urban environments is up to 50% higher than on highways, whereas one of the largest contributors to fuel use in urban areas is the stop-start behaviour of traffic flow. This phenomenon provides possibilities to address this area using ICT. Through the use of ICT, vehicles can communicate with the road infrastructure and other vehicles to obtain essential information to adjust driving behaviour. In recent simulation studies, using different times of preview information, it appeared that fuel savings could be achieved between 15 to 25% with 60 seconds preview and up to 33% with 180 seconds preview relative to an ‘un-intelligent’ baseline car. The development of a combined hybrid and ICTs equipped intelligent vehicle seems still under way with the optimal use of feedforward information as an ongoing research problem (Manzie et al., 2007).

3.2 Congestion Relief

Congestion is a function of the interaction vehicles have with each other due to speed flow relationships when volumes approach capacity (e.g., Black, 2003). It is not just vehicles going slow, or vehicles travelling at high speeds, that lead to congestion, although these contribute to the problem. The key to reducing congestion is controlling density, but this is not quite as easy as it sounds. Although the discussion thus far has focused on motor vehicles, the same observations could be made with regard to ships arriving at a port, or aircraft landing at an airport. There is a need to control the spacing (density) of these in a geographic or temporal sense.

Certain technologies that decrease the volume and density noted above will also go toward reducing congestion. The concern in this paper is for other types of ICT that will lessen the amount of congestion that takes place: Video Surveillance and Response, Informational Signing (variable messages), Advanced Traveler Information Systems, Adaptive Cruise Control, Intelligent Speed Adaptation, Congestion Free Zoning and Lanes, and Dedicated Short Range Communications. The ICT innovations that are fixed (or semi-fixed), i.e. Video Surveillance and Response, Informational Signing and the previously discussed Signalization belong to larger systems of Road Traffic Management that are currently in use in particular sections and nodes. In more advanced applications, mobile (in-vehicle) applications are being integrated with the fixed applications to arrive at a better fine-tuning of the systems and improve flow.
Video Surveillance and Response
Several cities maintain a continuous monitoring of key network locations to determine if traffic is moving or encountering congestion. Such monitoring can be done with strategically located sensors or television cameras. If flow interruptions are apparent they are usually caused by a disabled vehicle. Once these events are perceived, a repair/assistance vehicle is dispatched to the location. Upon arrival at the problem site, the objective is to remove the obstacle to flow and offer assistance (tire replacement, and so forth) or transport to the motorist.

Informational Signing (Variable Message Signs)
Electronic changeable message signs along the highway have proven to be of some assistance in communicating with drivers regarding major congestion points on the road ahead. Often these signs give directions as to ways to avoid upcoming congestion points related to accidents, congestion, and the like. It is important that such signs are not used on a continuous basis since drivers tend to ignore them if they always have the same type of message on them. A simulation study for different European city-regions on effectiveness of VMS (Variable Message Signs) on road network efficiency suggests quite modest results. Reductions in overall network travel times are 1-2% for the use of VMS in regular congested circumstances, provided that there is spare capacity in the network (Chatterjee and McDonald, 2004). Estimates for impacts on pollutant emissions and fuel consumption are similar to changes in overall travel time. Whereas the above changes are quite small, driver perceptions of the benefits turned out to be much higher. This points to a potentially important role for this application in the development of integrated transport strategies, because the provision of information may encourage the acceptance of demand management measures.

Advanced Traveller Information Systems
Personal information systems may take different forms and may be in-vehicle for car drivers and portable for passengers using public transport. Based on real-time information, the best route and connections (in public transport) are given. In advanced modes, opening times of facilities (shops, services, etc.) and the length of stays are used as an input, enabling an overall space-time optimization of activity chains. In the case of interruption (accident, congestion, etc.) new travel solutions are produced. Adoption of such traveller assistants - that are currently in an experimental stage - may be hampered by high costs but also by limited needs of travellers to plan their activity and their traffic chains. In a study of web-enabled information services on public transport carried out in England, Finland and the Netherlands, it was found that particularly the type of information provided matters (Molin and Timmermans, 2006). Real-time information turns out to be the most important attribute and this is followed by different planning options allowing search for routes that are not only shortest in time, but also cheapest and exclude interchanges. It appears that travellers are willing to pay for information under the condition that the information services provide added-value compared to existing information services.

Adaptive Cruise Control
Adaptive Cruise Control (ACC) is concerned with in-vehicle assistance to the driver in the longitudinal following control task. The main aim is to help to reduce congestion and smooth traffic flow, but an improvement of traffic safety is also hypothesized.
Experiments in Europe indicate that following with an ACC system can provide considerable reductions in the variation of acceleration compared to manual driving (40 to 50% reduction of standard deviation) (Marsden et al., 2001). However, this is true for long following sequences, whereas the results indicate that ACC systems may not be appropriate in those situations in which the driver needs most assistance, i.e. dense driving conditions.

**Intelligent Speed Adaptation (ISA)**
These systems also use in-vehicle electronic devices enabling one to automatically regulate vehicle speed. Like the previous technology, experiments indicate a higher effectiveness in less congested conditions (UK) (Liu and Tate, 2004). High speeds can be effectively suppressed, leading to a reduction of speed variation, but more slow moving traffic cannot be induced. In addition, it was found that ISA with full penetration could lead to a reduction of fuel consumption by 8%.

**Dedicated Short Range Communications**
These systems are based on information exchange between cars and may pertain to accidents, weather conditions, road construction, and similar events. In more comprehensive options, technical performance of the car can be communicated with the serving garage. Also, by integrating navigation systems, information about empty parking places and similar information can be transmitted to the driver. These systems partly rely on mobile communication between vehicles on the same route (oncoming and following traffic) and are still in the stage of development. “Early versions” are currently used in public transport (busses, taxi’s) and in freight transport.

### 3.3 Fatality Reduction

Road traffic accidents killed an estimated 1.2 million people in 1998 according to the World Health Organization (2004). More recent data are available for the EU 15 where there were 35,905 fatalities in 2003 (ERF, 2005). Comparable numbers for the US in 2003 were 42,643 fatalities (FHWA, 2005). For the most part and with some minor fluctuations, annual traffic fatalities have been falling in most of the developed nations of the world. At the same time forecasts continue to all for 1.2 million fatalities as far in the future as 2020 with decreases in the developed world more than compensated for by increases in the developing nations.

It is reasonable to examine what can be done technologically to improve the safety of road systems. Major improvements can be expected in two areas: vehicle safety and network safety. In terms of vehicle safety it is probably reasonable to expect a failsafe vehicle will be developed over the next decade or two. However, more important are those systems that improve vehicle-driving behavior. It is estimated that some 90% of all traffic accidents can be attributed to human failure, such as a lack of alertness or fatigue (Marchau et al., 2005). Vehicle radar technology is already available that warns drivers of obstacles in their path. The same technology could be tied into an on-board computer system and used to make it nearly impossible for the vehicle to crash into other vehicles or objects. It would do this by accelerating, decelerating, or stopping the vehicle. Today, advanced in-vehicles technology is available as options in up-market car models, like of Mercedes, Lexus and Citroën (NRC-Handelsbad, August 5, 2005).
Network safety seems to be heading primarily in the direction of automated guideways that would control the movement and speed of cars. Note that the technologies aimed at the previously discussed congestion relief sometimes also serve to reduce fatalities. Below, we address one “curative” approach, i.e. In-vehicle Accident Sensors and a range of accident “preventive” approaches, i.e. Vehicle Radar Warning, Blind Spot Information Systems and Night View Systems, and applications that serve both congestion relief and reduction of fatalities, i.e. Out-of-Vehicle Speed Control Systems, Advanced Drivers Assistance and Automated Guideways.

**In-Vehicle Accident Sensors and Radar**

It is generally recognized that many seriously injured individuals can survive such incidents if they can be transported to a medical facility quickly. The use of ICT in this case is intended to ensure this. A number of motor vehicle models being manufactured today come with sensors attached to the air bag system. Once the air bags are deployed, a communication of this event is sent to a dispatcher. The dispatcher in turn can communicate with the driver or other occupants of the vehicle and determines if any type of assistance (repair vehicles, ambulance, and so forth) is necessary. At this time several high-end models offer this service, however all models of General Motors in the US offer this. It should enable faster response to accident scenes than has been typical previously.

Numerous accidents occur when a vehicle in the process of moving in reverse hits a person or vehicle. Higher priced motor vehicles are now being produced that include radar installed in the back of the vehicle that alerts drivers of obstacles behind them. Mercedes (S-class) combines long-distance and short distance radar. The long-distance radar measures distance to cars in front of the driving car; the short-distance radar measures distance nearby in front of the car but also on both sides of the car. Of course, it will take quite some time before these systems are found in all vehicles of the fleet. In addition, there is also the problem of lower income drivers maintaining these and similar systems even if they are present (see Black, 2000).

**Blind Spot Information Systems and Night View Systems**

A number of accidents occur due to blind spots. To prevent such accidents, digital cameras are installed in the two outside mirrors that scan a zone on the sides of the car and produce a light signal if a car enters this zone. Volvo uses this system. Night View Systems, using infrared cameras are already available as an option in the most advanced models of Cadillac, Mercury, Lexus and Honda. These systems mainly serve to detect crossing passengers and animals during night. The problem is how to project the image without diverting attention of the driver from the road. Mercedes will install a small night-screen in the dashboard in the near future.

**Out-of-Vehicle Speed Control Systems**

Out-of-vehicle systems that control speed are quite commonly installed along roads in the US and EU and relate to adverse weather conditions and other incident conditions. A study in the US suggests that messages are significantly reducing speed in the area of adverse conditions, but that drivers tend to compensate for this reduction by increasing speeds downstream where such adverse conditions do not exist. Accordingly, this pattern casts some doubt on the net safety effects of speed advisory systems (Boyle and Mannering, 2004). A simulation study in Toronto (Canada) suggests that real-time
variable speed limits can reduce overall crash potentials by 5-17% (Lee et al., 2006). However, this study ignored the above-mentioned potential for compensation behaviour. In general, there seems a trade-off between the level of enforcement on driving behaviour and sustainability effects concerning emissions. A study in Portugal suggests that signal control schemes work differently for stopping cars compared to reducing speed of cars. Systems that stop a relatively large share of speed violators also yield higher pollutant emissions, whereas signals inducing speed reduction result in a decrease in relative emissions (Coelho et al., 2005).

**Advanced Drivers Assistance**

In the context of improving safety, we discuss the in-vehicle Automated Cruise Control (ACC) and Intelligent Speed Adaptation (ISA). Automated Cruise Control that primarily serves vehicle safety, performs both the longitudinal and lateral control task. Citroën today installs a system that warns the driver as soon as he/she moves to another lane without using the signal, by drawing attention through moving his/her seat. The lateral control task works by infrared sensors that measure variation in reflection of the standard markers on the road surface. In the EU, much research is currently devoted to in-vehicle collision avoidance based on sensor systems replacing infrastructure measures. An ultimate configuration is a $360^\circ$ car surround system as a “safety belt”. The systems that are currently studied vary in terms of technology, e.g. different radar sensors, infrared and visible spectrum imaging, laser technology, and in terms of distances and speeds involved (Lu et al., 2005). Research into such systems is in progress today, but the systems are still in an experimental stage waiting for solutions that are more robust, i.e., not vulnerable to influence of weather/atmospheric conditions and interference with other electronic systems, and more acceptable in cost or price. For example, Lexus plans to introduce lighting systems that monitor speed, braking performance and weather conditions, and automatically adjust the amount and type of lighting as a warning (active lighting).

Quite some attention has been paid to the impacts of Intelligent Speed Assistance (ISA). We mention estimated safety effects of full automatic speed control devices up to a 40% reduction of injury accidents and 60% reduction of fatal accidents (e.g. Marchau et al., 2005). For the Netherlands, estimates reveal a fatality and heavy injury reduction of up to 30 and 38% on roads with speed limits up to 90km/h. In addition, estimates of the impact of automatic positioning and collision avoidance systems indicate similar maximum reduction levels for particular systems on particular types of roads.

A special category of in-car safety systems are brake control systems, e.g. working through an alarm. New in this respect is the “intelligent brake control” that becomes activated as soon as the driver shows a panic reaction (release of pedal). It prepares the brake control in such a way that all braking power becomes available as soon as the driver puts on the brake. If necessary, electronics takes over mechanical power in brake control because it reacts quicker and more refined in terms of using the right braking pressure.

**Automated Guideways**

Ultimate network safety can be reached with automated guideways. Test facilities of automated guideways have been developed by Honda and Nissan. One could pull the vehicle onto such a guideway and the system would take over control of the vehicle. A somewhat related idea would have vehicles linked electronically if they were traveling
to the same destination. Such bonding would probably be possible on existing roadways and would increase vehicle density without necessarily decreasing speed. Of course, legal issues concerning responsibility and liability between drivers, network operators and ICT system manufacturers are quite different from conventional driving and need to be settled.

4. Evaluation

The above discussion suggests that ICT innovations in the transport system seem to be most effective in fatality reduction. The literature gives an estimated reduction of fatalities up to 30 and 40% and even 60% by particular types of Advanced Drivers’ Assistance in particular sections of road networks. What seems a fortunate situation is that Advanced Drivers’ Assistance at the same time may serve congestion relief, although the results seem much less convincing than in fatality reduction. At the same time, most recent research suggests fairly high potentials of fuel-intelligent cars that ‘communicate’ with the road infrastructure and other vehicles to adjust driving behaviour in optimizing fuel use, i.e. a reduction by approximately 30%.

In general, there are three constraints for adoption of new ICT applications in the urban transport system, i.e., (1) high costs for users compared with perceived benefits, with user costs increasing if more than one layer in the system is involved, including fixed infrastructure, (2) technological and actor complexity, similarly if more than one layer is involved, and (3) legal issues that are not sufficiently settled, e.g. concerning responsibility and liability of the actors involved in the case of failure of the new application (e.g. drivers, manufacturers and operators of the ICT systems). Taking these factors into account, it seems that systems of Automated Guideways, either fixed or electronically, face the smallest potentials to be adopted in the short and medium term. Based on the above factors, it seems plausible that potentials for adoption in various countries and urban regions may differ according to various circumstances (Black and Van Geenhuizen, 2006):

− driving circumstances, e.g., larger or smaller distances, a different occurrence of high-density urban areas leading to different levels of annoyance of road congestion, and different maximum speed levels on particular types of urban roads;
− needs for car driving, e.g., automobile use to satisfy needs for ‘driving experience’ in relation to public transport use;
− car cultures, e.g., cars may be more or less a symbol of freedom and status;
− institutions, e.g., a weak or strong government involvement in transport, a weak or strong focus on legal liability issues, different levels of taxation of car ownership and use, and different systems of incentives in promoting sustainable transport.

The above factors indicate quite some differences like between the US and countries in Europe.
5. Conclusion: quite some ignorance

The above discussion on impacts of ICT use on the working of the transport system provides ground for the following observations. There is still some ignorance about the sustainability impacts of various ICT innovations in transport, (e.g., Van Geenhuizen and Thissen, 2006). This situation is mainly caused by:

- A fragmented character of the research that has been done; exceptions are the impacts of ICT use on fatality reduction, particularly off-vehicle speed limitation and in-vehicle driver assistance; fragmentation in research is caused by a widely varying interest of market parties to push new ICT applications.
- Simulation-based research without sufficient small-scale real-life experiments and large-scale research on real travel behaviour that can increase the validity of results.
- A limited scope of much research, namely confined to particular parts of the transport network; as a result there is an ignoring of potential compensation behaviour elsewhere in the network and of potentially secondary adverse effects like relocation of households at a larger distance from work due to a more efficient commuting.

It is clear that additional research is needed to fill the above indicated knowledge gaps, particularly the question whether significant improvements in particular parts of the transport network go along with sufficient overall network performance.

ICT innovations seem to be most effective in fatality reduction. The literature mentions reductions on a level of 50 to 60% and on a level of 30 to 40% for particular types of advanced divers’ assistance in particular road network sections. Measures to relieve congestion seem much less effective and work only under restricted conditions, namely spare network capacity and lower levels of congestion (density). The latter suggests that the ICT innovations concerned cannot yet fully work under conditions that they aim to solve, that is under high congestion levels. This calls for further research to identify ICT technology that is particularly effective in such conditions, or that will improve the existing applications that are currently less effective. Policy efforts, for example to increase R&D on particular ICT applications or to make particular applications less expensive for users, need to focus on those applications that don’t suffer from technological uncertainty and from uncertainty due to legal aspects or inherent actor complexity. Accordingly, it seems more realistic to promote the adoption of a smart set of relatively simple applications than the adoption of comprehensive systems.

Acknowledgements

This paper is an improved version of a paper presented at the IX NECTAR Conference, Porto, Portugal, 9-12 May 2007.

The study has benefited from financial support by the Delft Research Center of Sustainable Urban Areas, Delft University of Technology.
References

Research Outlook on a Mixed Model Transportation Network

Joakim Kalantari\textsuperscript{1*}, Henrik Sternberg

\textsuperscript{1} Division of Logistics and Transportation, Chalmers University of Technology, Göteborg, Sweden

Abstract

The purpose of this paper is to present an outlook on the future research for a specific mixed model transportation network referred to as Foliated Transportation Networks (FTN). FTN is thus far a conceptual model that is based on the idea of foliating a direct shipment and a hub-and-spoke structure in order to achieve higher fill rates without an increase in the total traffic work at the same time. The conclusion is that the two principal areas of research are areas of planability (i.e., the ability to in advance and on a sufficient level of detail and precision determine the capacity requirements of the system) and network optimization (i.e., the optimization of the distribution of goods and resources between the different layers of the network).

Keywords: Foliated transportation networks, Mixed model transportation, Transport, Transport planning and control, Transport network optimization.

1. Introduction

Looking at the road bound freight transportation industry; one will quickly encounter a system wide overcapacity and low utilization rates regarding load capacity of the transportation units, i.e., low technical efficiency. Previous research indicates a normal system fill rate span of 50-70\% (Caputo et al., 2005, Nanos-Pino et al., 2005).

The technical inefficiencies in the road bound transportation systems have significant negative external effects, making efficient use of the physical resources an interest for society as a whole. Trucks and trailers impact the environment negatively both in terms of pollution (e.g., \(\text{CO}_2\) emissions, air and water pollution, etc.) and in terms of noise pollution, congestion and traffic hazards (Kreutzberger et al., 2003, McKinnon, 1994).

The steady past and projected growth of the transportation demand (European Commission, 2003) amplifies the significance of this issue especially in light of the fact

* Corresponding author: Joakim Kalantari (joakim.kalantari@chalmers.se)
that historically, the growth has not been absorbed by the existing overcapacity (Rodrigue, 1999). This observation suggests that the transportation system structures are constructed in a way so that they require or are benefited by maintaining an extensive overcapacity. This development is critically alarming because what the resource overcapacity aims to satisfy is being disabled due to the predictable near saturation of the infrastructural capacity (Crainic et al., 2004).

Efficiency of transport networks is dependent on effective planning (Acharajee, 2000, Landers et al., 2000, Closs et al., 2005) and the alarming background calls for new ways of operating transport networks. Considering the substantial improvement potential that exists, due to the low utilization rates this matter is highly interesting from both societal and industrial perspectives. Changing cost structures due to the rapid development of energy prices and political sanctions as an attempt to dampen the transport sector’s negative environmental impact also heavily contribute to the growing need for increased efficiency along with improved or at least sustained performance (McKinnon, 1995, McKinnon and Ge, 2006).

A conceptual model that attempts to address these concerns is Foliated Transportation Networks (FTN). The purpose of the concept is to improve the technical efficiency of the transportation system, i.e., resource utilization regarding loading capacity without the deterioration of other quality or performance parameters. The conceptual model of foliated transportation networks (FTN) is a hybrid model that aims to improve the efficiency of a transportation system by combining the two predominant network structures, i.e., direct shipment (DS) and hub and spoke (HS). It is stipulated that by foliating the two structures (i.e., DS and HS), and by dynamically planning, controlling and optimizing the distribution of goods and resources between the two sub structures, strengths of the individual setup will be amplified at the same time as their weaknesses diminish, resulting in better system performance than any single one on its own (Persson, 2006b, Persson and Waidringer, 2006). In other words, the physical resource utilization will be increased without the deterioration of the other key performance measures such as total traffic work, number of resources in the system, lead time, flexibility, etc.

The purpose of this paper is to presents a research road map for developing the concept of FTN to an operational model. The road map contains both an overview of the empirical as well as theoretical gaps that need to be filled in order to establish the concept of FTN. The contribution of this paper outlines the current status of the research and identifies the necessary research areas for accomplishing an implementation of FTN. The research on FTN takes the perspective of the transportation service provider and regards general cargo freight.

Based on the description above, several principal problems will be posed in contrast to the existing systems. These issues need necessarily be identified and resolved before it is feasible to consider the model operational. This paper aims to identify these gaps.

---

1 Going forward, “the hypothetical model of foliated transportation network” will simply be referred to as FTN or variations of the same.
2. Methodology

Preparation of this paper involved a literature study and collection of empirical data. The literature study uses secondary sources such as books, the Internet and scientific articles within the areas of logistics and transportation, information science, mathematics and transportation planning and control. The empirical data employed is qualitative and has been collected via semi-structured interviews and observation. Two major terminals have been visited and personnel from senior management to individual operators have been interviewed. The interviewees have been selected through a process of “snowballing”, where each new interview reveals the need, identifies the interviewee and creates access to the next one.

This research takes an exploratory approach to outline the research outlook for FTN. This is done by systematic combining, i.e., combining inductive and deductive methods (Dubois and Gadde, 2002).

The process of identifying and analyzing researchable gaps for the FTN considered the following factors:

− No real world implementation of the concept available
− The inherent complexity of the studied phenomena (transport networks) (Manheim, 1979, Nilsson, 2005, Waidringer, 2001)
− Sparse literature on the concept (Persson and Lumsden 2006)
− The interdisciplinary nature of the concept, e.g., logistics and transportation, mathematics, operations research and informatics (Persson and Waidringer, 2006)

These factors led to an iterative process, consisting of the three interrelated research components: theory (literature study), empirical area (empirical data collection) and researchable gaps (analysis). This approach is similar to the so called “whirl-pool approach,” which has been successfully applied in areas such as computer science and information systems (Williams, 1996, Travisano, 1996).
The problems from the empirical area generated the initial literature studies. The concept applied to the empirical area generated researchable gaps, which in turn gave indications for additional theoretical areas. As illustrated in (Figure 1), several iterations were carried out before the researchable gaps and were summarized and categorized into the resulting research outlook.

Semi-structured data collection interviews have been performed with senior managers, sales representatives, information system officers, terminal managers and operations managers. Floor operatives, planners and individual haulers have also been interviewed. The need for additional data gradually arose during the iterative course of consulting the existing theory, collecting empirical evidence and submitting the thus-far results of the analysis to formal evaluating seminars. This process was repeated until saturation was reached (Flick, 2006).

The Swedish domestic general cargo freight transportation market is dominated by two large companies, each covering the entire country through their own network (Sommar and Woxenius, 2007). This considerably narrows the choice of where to collect empirical data, seeing how FTN is only applicable for large networks carrying an abundance of cargo. In the choice of which company’s transport network to use as empirical starting point, the most important factor was that the studied network should meet the representative rationale (Yin, 2003). The objective of the representative case is to capture the circumstances and conditions of an everyday or commonplace situation. According to Yin (2003): “The case study may represent a typical project among many different projects, a manufacturing firm believed to be typical of many other manufacturing firms in the same industry, a typical urban neighborhood, or a representative school, as examples. The lessons learned from these cases are assumed to be informative about the experiences of the average person or institution.”

The actual company chosen has been used by other studies for validation purposes (Stefansson, 2006, Sternberg, 2008a), which is a factor supporting the representativeness of the empirical data. For a full account, please refer to the reference case description.

In order to bring additional validity to the results of this study, the identified research gaps have been discussed, revised and finalized through a series of eight formal seminars. The participants of these seminars have been transportation practitioners and research professionals as well as researchers of other neighboring disciplines such as traffic, logistics and supply chain management. The results, presented in this paper, have been finalized after being put to academic scrutiny i.e. defended against a senior and a junior opponent, in a final seminar.

3. Theory of FTN and related areas

The purpose of the theory chapter is to put the concept into a transportation context, outline previous research, describe related concepts and present the theories related to the gaps identified.
3.1. Transport operations management

Transportation is normally associated with the movement of a product from one node in the distribution channel to another. In transportation, attempts are made to solve this problem by ensuring that products are moved as quickly, cost-effectively and consistently as possible from the point of origin to the point of consumption (Ross, 1996).

The transportation network consists of nodes and physical links. Links correspond to highways, rail tracks, seaways or urban streets and nodes express the connectivity relations of links in the network, e.g., warehouses, distribution centers, freight terminals and ports (Manheim, 1979). There are many individuals, groups and organizations whose decisions interact to affect the transportation system and thus the pattern of flows. The user of the transportation, i.e., a shipper of goods, makes a decision about when, where, and whether the goods should be transported and how often (Manheim, 1979).

Options or decision variables are those aspects of the transportation system that can be directly changed by the decisions of one or several individuals or organizations. Manheim (1979) outlines possible options when it comes to transportation system operating policies. This set of options includes the full spectrum of decisions about how the transportation system is operated. The networks, links, and vehicles establish an envelope of possibilities; within that envelope a large variety of detailed operating decisions must be made. These options include vehicle routes and schedules, types of services to be offered, including services auxiliary to transportation (diversion and reconsignment privileges for freight), and regulatory decisions. A transportation setup is a set of decisions based on options available for a certain flow of goods.

In general, transport operations management concerns five corner stones and their interrelations. These five are (Brehmer, 1999):
- the capacity of the studied system and its resources
- the average time required for each activity
- the coordination of the resources and the activities
- the inventory, and finally
- the control of activities

A key enabler of effective transport operations management is having necessary information available (Landers et al., 2000). Every setup requires exchange of information between all the involved participants in order to avoid execution hurdles (Sternberg, 2008a).

A good shipped in the system is typically heterogeneous, i.e., loading and unloading procedures varies from shipment to shipment. Arnäs (2007) defines heterogeneity as: “… goods where one or more of the parameters Density (D), Stowability (W), Ease of handling (H) or Liability (L) are outside of their accepted range.” In this case it means that a majority of the goods is not fit to be placed, or simply is not placed, on standardized load units or the goods do not offer standardized interfaces for handling.

3.2 Foliated transportation networks

In a transportation network based on the principle of direct shipment every terminal (i.e., node) in the network is connected to all the other ones via direct relations (i.e., link) (Manheim, 1979). This means that dedicated trucks traffic the relation without any
other consolidation stops. In an (from the operator perspective) ideal setup, there will exist a set latest time for accepting new orders. Placed orders will be collected by milk-runs and delivered to the nearest terminal for consolidation and transshipment. All the goods with the same destination will be sorted out and loaded onto long haul trucks in order to be delivered to the destination terminal where they will be sorted and distributed to their final destination, i.e., to the individual consignees.

In contrast to the DS setup where every origin/destination pair of terminals are interconnected through direct relation, in a HS setup all the terminals only have a direct relation to the/a hub. A network can consist of a single hub where all the terminals are connected or multiple hubs where every hub has a direct relation with a cluster of terminals in its vicinity and all the other hubs in the system. The difference between a single and multi-hub setup is the additional consolidation/deconsolidation step that the inter-hub transports imply. In the following presentation a single hub setup is assumed for reasons of simplicity.

When the goods are collected from the consigners, in a HS setup, and delivered to the origin terminal the consolidation process is performed irrespective of the final destination. All the goods from every terminal are then delivered to the hub, where the goods are sorted according to their destination.

However, generally DS and HS are considered each other’s opposites. Where DS affords quick and easy to control transports with minimum transport work at the cost of the system wide fill rate of the trucks, the increased number of trucks, and dependence on large volumes of goods in every direction (Crainic, 2002), HS is the exact opposite (Kalantari and Lumsden, 2007). In a HS setup the goods never travel the shortest way, yielding a longer transit time, more handling and more transport work, instead resulting in fewer trucks necessary, higher fill rates and less dependence on big volumes in the system (Kalantari and Lumsden, 2007, Lumsden et al., 1999).

The principle of FTN, where the two structures are foliated, is meant to roughly lead to the following procedure/structure. Every terminal in this setup is interconnected with direct relations. This implies of course that the ability for direct deliveries and hub consolidations are not restricted. After the initial collection milk-run, all the goods with the same destination that collectively fill whole units (i.e., trucks or semitrailers, etc.) are sent through direct relations i.e. according to the DS principle. All the remaining goods (i.e., goods that do not have the sufficient combined volume to fill a whole unit in a DS relation), in this context defined as “overhang” (Persson, 2006a), are sent to the hub for consolidation (see Figure 2). The conceptual assertion is that the aggregation at the hub will contribute to first diminishing some of the negative traits of DS as well as improving the physical performance by introducing the positive qualities of HS on the remaining volumes or the overhang (Persson and Lumsden 2006, Persson and Waidringer, 2006). When viewed in isolation, the different categories of goods are theoretically well suited for the different network layers that they are meant to travel through, which gives the concept some additional strength (Hultkrantz, 1999, Kalantari and Lumsden, 2007).

---

2 This, of course, assumes that all terminals are unrestricted in capacity. In reality, not every terminal is suited to be selected as a hub, both due to capacity constraints and geographical location.
3.3. Related network concepts

What separates the idea of FTN from other concepts and solutions now in practice or in the focus of the transportation and logistics research community is the aim of foliating two different networks structures that are to be managed through dynamic and systematic planning and control (Persson, 2006a, Persson and Lumsden 2006).

Some of the other solutions that exist and are in operation lack some of the characteristics described above in that they are managed in an ad hoc manner. Furthermore, none seem to strive to deal with this issue systematically or at the system level. Some examples are stand-by shipments, priority classes of goods and price differentiation. Analogies can also be drawn from the field of air transportation, e.g., air passenger and/or cargo transportation.

Stand-by shipments refer to goods that are accepted and stored at the terminal and will only be delivered when overcapacity arises, i.e., the buyer accepts a flexible lead time. In practice, those types of goods will be loaded any time overcapacity in the required relation occurs; for instance, cars at Scandia Harbor\(^3\) or tires at Schenker’s Bäckebo Terminal.\(^4\) These examples elucidate both the ad hoc nature of the solution and the lack of system overview (i.e., only the selected long distance legs are regarded on stochastic bases). Different priority goods are based on the same basic principle as the stand-by shipments, where for a lower price the customer allows the goods a longer transit time; for example, the free delivery option of books from Adlibris\(^5\), an online book store. This access to less time sensitive goods allows the transport network operator to distribute the low priority goods so as to make use of the overcapacity. Here the entire system can be affected. However, the approach still lacks systematic and dynamic planning and

\(^{3}\) Volvo Cars utilizes the over-capacity of the transatlantic ships for deliveries to America. The cars are temporarily stored adjacent to the port in question.

\(^{4}\) Michelin Tires are temporarily stored at the Bäckebo terminal and are sent across Sweden using the overcapacity of Schenker’s long haul trucks.

\(^{5}\) When ordering at the online book store Adlibris, one gets to select a normal delivery (1-2 days) that will require a fee or to allow the shipment to arrive within 2-5 days for no charge. Naturally it is the service of the transportation provider that is interesting here, and not the pricing strategy of the online bookstore.
control. The third example is more a strategy for coming to terms with flow imbalances between regions and O/D pairs. The basic idea here is to generate new flows from old destinations. Another common difference between FTN and the examples is the fact that the excess capacity is undervalued in the examples, whereas in the FTN this is not meant to be the case.

Looking at the air transportation networks, similarities between FTN and passenger/freight transportation can be identified. For passengers, the basic idea of FTN is implemented in all large networks with one crucial difference. In a network of air transportation for passengers, the detailed capacity planning is not a fundamental part of the equation as passengers opt for different routes according to availability or price parameters at the time of booking/purchase. Also, the cargo in the belly hold of an airliner is accepted on grounds similar to that of a stand-by shipment with almost exactly the same effect (Acharajee, 2000).

Besides the difference in the level of planning (ad hoc vs. systematic and dynamic) and the level of attention (system wide vs. specific relations/units/origins/destinations) the FTN is distinguished from its predecessors/family members in that it aims to foliate two different network structures. This is an effort, the success of which would in principle be interesting even in other contexts besides the current one, i.e., the Swedish domestic road bound general cargo freight transportation. The concept of FTN, as presented here, has not been developed beyond its conceptual cradle and in its goal state, i.e., a fully operational model is unique and novel with a promise result that makes it well worth the research effort (Persson, 2006a, Persson, 2006b).

3.4. Optimization related to transport network

A special bounded case of the general 0/1 knapsack problem where the weight and profit are equal is called the subset sum problem (Kellerer et al., 2004, Martello and Toth, 1990). One way of looking at a Bin Packing Problem is a special case of a multiple subset sum problem where the capacity is constant (Martello and Toth, 1990). A Bin Packing Problem is a combinatorial problem of maximizing the use of a limited discreet resource and is referred to by various different names throughout the literature, e.g., cutting stock or trim loss problem, bin or strip packing problem, nesting problem, etc. (Dyckhoff, 1990).

The packing and cutting problems up to three dimensions have an easy to recognize connection to their physical applications (the general problem is not limited to three dimensions). For instance, the one dimensional cutting could be cutting as in cutting pipes or logs (Dyckhoff, 1990), the two dimensional could be cutting shapes out of sheets of paper or cloth (Lodi et al., 2002) and the three dimensional cutting could be filling a transportation unit or bin (Silvano et al., 2000). These examples are by no means exclusive or exhaustive. The mathematical problem could pertain to any planning or scheduling problem to fit the description above (Chantzara and Anagnostou, 2006, Scholl et al., 1996). As the term Bin Packing Problem suggests, one of the primary applications lay within transportation (Dyckhoff, 1990, Gehring et al., 1990, Silvano et al., 2000).

Another particular bounded case of the general knapsack problem is the so-called Change Making Problem where the profit is constant and set to 1 and the capacity is constraint (Kellerer et al., 2004, Martello and Toth, 1990). This issue is connected to
general cargo freight transportation in the sense of maximizing the number of shipments served by one loading unit.

4. Reference case

The reference case provides the empirical background or current state of affairs for the analysis. The studied national network, belonging to a global logistics and transportation provider, consists of a network of over 30 terminals, covering virtually every corner of Sweden. The goods shipped by the provider are divided into two groups: general cargo, which is handled through terminal(s) and is of specific interest for this study of FTN, and direct goods which are constituted of large shipments that do not pass through any terminal. Routinely, consignments totaling a combined equivalent weight less than 1000 kg are considered general cargo, regardless of their shape and packaging. The focus of this study is the general cargo section of the national network. This part of the system handles roughly 5000 tons of freight daily.

A very limited number of the trucks are actually owned and operated by the company which instead makes use of risk and profit sharing programs with “independent” haulers. The word independent is within quotation marks because the majority of haulers are in fact not independent in the true sense of the word due to the terms of the contract. Based on mutual concessions some rights and duties are shared between the company and its contracted haulers, e.g., a hauler may have the exclusive rights of a line and in return shares the risk of not getting paid for unutilized capacity.

The information system is fragmented and different sub-sections are unable of communicating with each other in real-time or automatically. For instance, the capacity and profit sharing system, the goods information system containing EDI (or physical) waybill information and the billing systems are isolated from one another. Counter intuitively, this is an advantage for this decentralized system as the experienced based low level control actually outperforms any existing control system in this highly peculiar, fragmented and low level autonomous setting. This is actually considered a positive trait of this company as a case to be selected for the study of FTN, because the system is comparable to a network of cooperating haulers with no uniform information system, which would have to be the case on the European continent where no one provider has such a dominating position in the market. This in turn could expand the scope of generalization within the boundaries of what is possible from a single case study (Yin, 2003), regarding the results.

Over 50% of the available capacity in the general cargo system is pre-booked. This means that at any given terminal on the basis of experience, implicit and explicit customer contacts, a certain amount of capacity is booked for some of the customers where they will report deviation from the norm, instead of booking, even in a case where no formal agreement exists. At any rate, pre-booked or not, more than 85% of the customers send the waybill information regarding their goods via EDI.

This information, though readily available in the system, is not utilized for planning or control. Even though the identification numbers of the physical waybills are scanned

---

6 This of course does not apply to goods requiring special attention, i.e., chemicals, provisions etc. that are bound by law to be treated separately.
2-5 times during the door-to-door service, they are not automatically connected to the corresponding digital information regarding the consignment. They can, however, be matched manually after the fact. The point of including this information is to clarify the decentralized nature of the control of these operations and the lack of central and formalized planning and control. This also furthers the analogy of comparing this single company to a network of intercompany cooperation.

A final note requiring explicit attention is the fact that the general cargo transported through this network is heterogeneous at best.

5. Analysis

On the basis of the framework and the reference case, the analysis outlines the identified research gaps following the order in which they were identified. A holistic approach is taken, which takes into consideration the added complexity of multiple unsolved issues.

5.1. Load capacity planning and control

One of the assumed prerequisites of FTN is improved or at least sustained performance compared to the current setup. This notion is the basis for the first and potentially the most critical suite of concerns.

As touched on earlier, the traits of the DS and HS setups admit that the HS setup requires more transport time than the DS structure. However, in the description above, it is apparent that what constitutes “the overhang,” i.e., the goods that are to travel through the HS layer of the network, are only available after all the DS shipments have departed. This assumes that the same control and planning principle that is used today is to be utilized even in the FTN setup. Simply put, the goods that are to travel the furthest total distance and require extra handling, i.e., extra time in transit, will depart last. This expansion of the time in transit is likely to deteriorate the performance of the system either because the goods will arrive later or orders must be placed earlier than in the current system.

One viable approach for resolving this issue is shipping, in every relation (origin destination pair), the portion of the goods that is to be the overhang first. This measure would allow the most time consuming activity the largest amount of time. This approach, however, will require an unprecedented level of capacity planning detail (Sternberg, 2008b, Christiaanse and Kumar, 2000). That detailed level is difficult to achieve for homogeneous goods in the current systems, not to mention the implications of attempting to do so regarding heterogeneous goods.

Exploring this path, many other challenges will need settling. Aside from the obvious question of the feasible level of planning and control that is to be achieved, other questions intimately associated with this issue will also need to be scrutinized. The quality and level of detail in the current customer orders needs to be established and compared to what level and quality of order data is absolutely necessary, preferable and/or feasible to expect from the transport buyers. Furthermore, the question of the impact of the need for information and accuracy, and the methods to obtain those, on the flexibility and robustness of the system, deserves attention. This also relates to
another question, the level of rigor of the governing rules of the system e.g. latest time of order entry, tolerance for accommodating last minute changes, the parameters for decentralized decision making and the ability of the system to cope with deviation from the plan need to be determined (Sternberg, 2008a).

5.2. Adding the dimension of heterogeneity

The aim of the central planning and control operations in this context is not only to prevent capacity shortages but to ensure the most efficient use of the capacity available. This notion is at the very core of FTN. The ability to effectively plan at a sufficient level of detail is one of the chief prerequisites for obtaining an efficient use of the physical resources and is thereby a system with superior performance (Power, 2005). In fact, one could argue that inept capacity planning could result in a scenario where the gains from the foliated structures would be outweighed by the poor utilization that could be a direct result of an inadequate planning effort. This respect of the issue and the impact of its outcome are amplified because of the fact that the “overhang” is to be, in advance and on a great level of detail and accuracy, identified and shipped from the terminal first.

The heterogeneity of the physical properties of the goods leads to a setting where the loading composition and loading sequence of the goods within the transportation unit will, to a degree not quantified at this moment, affect the loading capacity required. This means that even if the physical utilization rate regarding, e.g. volume, weight or length could be measured precisely (which may not even be the case) the reverse relation, i.e., planning for capacity based on aggregated weight, volume or length parameters cannot be taken for granted. It is imperative to establish what is possible to achieve in terms of planning precision for heterogeneous goods based on the existing order data. The inquiry ought not ignore what input will be required to achieve sufficient planning detail and whether other factors can be entered in, which could ultimately result in enhanced planning performance.

5.3. Adding the dimension of the Bin Packing Problem

It is not, then, only a matter of information quality/requirements on its own, but one that is complicated with a multi-dimension, multi-choice, multi-constraint Bin Packing Problem. Crassly, it means that given perfect information and absolute ability to simulate the operations necessary, the issue is still unsolved. The Bin Pack Problem, i.e., how to best fill the loading units to achieve optimum utilization, remains to be settled.

The problem is, strictly mathematically speaking, non-trivial and stands unsolved at the moment. It would be much more interesting to continue the research in this area in light of a mathematical solution to these specific problems. This also implies that the company specific traits that might add to the level of difficulty probably won’t significantly alter the principle solution.

5.4. Adding the dimension of handling speed

A more, seemingly, trivial matter concerning the time aspect is one of reducing the total time in transit for the goods within the HS layer of the FTN. In order to keep that time to an absolute minimum, strategies and procedures must be devised so as to ensure
speedy handling through the additional consolidation point/s, i.e., the hub/s. The number of hubs and the extent of the detour that a hub transshipment entails can considerably affect the additional time in transit required (Kalantari and Lumsden, 2007). Therefore, in some instances it may be necessary to reduce the handling time, while on other occasions this may just improve the performance. Several approaches have been pointed out in the past, e.g., only selecting easily handled goods for the HS layer (Bjeljac and Dalibor, 2004), pre-sorting in cages (Acharajee, 2000), use of RFID (Persson, 2006a) or any combination of these and other possible operational solutions. The question of which approaches are necessary and/or feasible and to what extent, needs to be addressed.

5.5. Adding the dimension of a Change Making Problem

Mathematically, here also lies a non-trivial Change Making Problem to be addressed. This Change Making Problem, though non-trivial and unsolved in the general form, will need qualitative inputs from the refined system design for its specific solution. These inputs need to be identified and defined for a specific solution to be feasible. For illustration, the profit variable in the Change Making Problem could be as easily handled as the number of shipments put through the hub (minimize function) or a compound measure to be defined later in the lines of “handleability,” i.e., ease of terminal handling (Lumsden, 2006).

5.6. In summary of planability

This portion of the complex of problems presented above can be referred to as planability, i.e., the level and detail of planning and control that is required/feasible and its possible consequences.

In summary, it is not only a matter of data quality but it also concerns the magnifying level of the data (i.e., the level of detail and richness of data) in combination with the specific requirements of heterogeneous goods and the knapsack problem, i.e., the ability to successfully transform the data into a solid plan. The tricky part is that even success in doing so does not clear the fog and automatically lead to an uncomplicated state of affairs. There will be a point of balance between where central planning will enable better resource utilization and where it will inhibit flexibility and robustness. In order to be able to make a business decision about this trade-off, this interrelation needs to be cleared and to the extent possible, quantified. The combined effect of the planning errors, i.e., the knapsack problems and the error of the available/feasible information obtained from the consigner, like inadequate or faulty information, last minute changes etc. need to be examined.

5.7. Global Optimization

While the matters discussed in the previous sections are some of the key factors to operationalise the conceptual model of FTN, the following section is devoted to maximizing the potential of the same in practice. This area of inquiry is considered secondary and must be based on the outcome of the planability issues. Before continuing any further it must be stated that the term “optimization” is in fact incorrect, strictly speaking. As it will be apparent below, it will probably be too difficult or even
impossible to find an optimum. A sufficiently improved solution or a near optimum will for practical reasons be what is sought.

5.8. Governing rule issues

As repeated several times before, the goods that travel through the HS layer of the system will not only be in transit longer than the ones in the DS layer, but they will travel a greater distance as well. The point being made is that by sending a slightly less than 100% full truck into the DS layer, the extra distance and handling for that entire almost-full truck will be eliminated compared to it travelling though the HS layer. This then begs the question: What is the filling rate of a “full” truck? Could that be situation dependent? Or, is there a set estimated value to follow? How substantial is the effect of these decisions on the system wide performance?

Another aspect is that of hub capacity. Assuming a truck will have a fill rate slightly below what has been established, dynamically or otherwise, as a full truck, putting that additional truck through the hub would seriously congest or deteriorate the hub’s ability to handle the extra volume. Another problematic area is the impact delivering an additional truck to the hub will have on the trucks leaving the hub, i.e., the resource requirements of the hub-destination relation will be affected. For example: the system wide fill rate of a perfectly balanced hub with as much goods coming in as going out and a 100% fill rate for all leaving trucks might suffer immensely if an extra truck load were to be added to that balance. How could these aspects be considered when deciding where to send the goods systematically and automatically? How does this change the application of the solution of the aforementioned knapsack problems? These would be other factors that would require special attention when deciding the refined governing rules of a FTN. Even though these concerns are rooted partially in qualitative concerns affecting the governing rules of the FTN some equally vital mathematical issues need attention as well.

The heterogeneity of the goods contributes additional complexity to the problem. A clear notion of the solution of the knapsack problem combined with the issue of information quality is necessary for it to be possible to proceed with the optimization question. This is a question which on its own, without the complicating aspects injected, is certainly complex enough.

Clearly, these types of problems cannot at the moment be solved through analytical mathematics. Other solutions ought to be pursued, such as rule-of-thumb or a heuristics based decision support system. The two question areas posed above sum up to a system where, in theory, the distribution of goods within the different sub structures of the system needs to be reconfigured after every order entry. An alternative to the abovementioned approaches could be the exploitation of ideas from adaptive logistics (Dorer and Calisti, 2005, Nilsson, 2005) and agent based modeling and simulation (Neagu et al., 2006). These speculations will, however, not be further developed in this document since they fall outside of the scope of this paper.

The aspect that is closely connected to this area and its inquiry and should perhaps precede the optimization research is the quantified theoretical and feasible practical potential of a functioning FTN. The rules governing such an unpolished still “under construction” system could in a sense be arbitrary, as long as they adhere to the basic conceptual idea.
6. Results

The results of the analysis of research gaps related to FTN are illustrated in figure 3.

![Figure 3: Road map of FTN research. Circles depict areas of science, boxes depict researchable gaps and the arrows depict the suggested research path.](image)

Both literature and interviews identified the load capacity planning and control (transport planning) as the initial research point of the concept. Transport planning lies in the intersection of transportation and information science. Adding heterogeneity to the transport planning creates new dimensions to all of the planning related areas, both to mathematical optimization and information requirements to operationalise FTN. These gaps find a relevant formulation and possible solution in the literature of combinatorial mathematics in general and specifically as general problems within the Bin Packing Problem and Change Making Problem. The transportation planning and control required are concluded to be dependent on adequate solutions to correctly formulated specific cases of these general mathematic problems.

Through the workshops and interviews, and also based on previous research, the handling speed was identified as another important issue to deal with in the design of an operational model. Finally, several other issues related to the control of the system were
identified, labeled as *Governing rule issues*. The current state and level of abstraction of this conceptual model leaves plenty of unresolved questions regarding these aspects, which are closer to an operational design of the model.

### Table 1: Research gaps and road map to future research on FTN.

<table>
<thead>
<tr>
<th>Research Gap</th>
<th>Research area</th>
<th>Implications</th>
<th>Relation to FTN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load capacity planning and control</td>
<td>Transport planning Information systems</td>
<td>Derived mainly from study on management in the empirical setup and transport literature</td>
<td>The core of the operational model</td>
</tr>
<tr>
<td>Heterogeneity</td>
<td>Transport planning Information systems</td>
<td>Mainly derived from the empirical setup</td>
<td>Needs to be addressed in order to make the model operational in a large-scale setup</td>
</tr>
<tr>
<td>Bin Packing Problem</td>
<td>Transportation Mathematics</td>
<td>Theoretically derived</td>
<td>Planning and control algorithms</td>
</tr>
<tr>
<td>Change Making Problem</td>
<td>Transportation Mathematics</td>
<td>Theoretically derived</td>
<td>Optimization algorithm</td>
</tr>
<tr>
<td>Handling speed</td>
<td>Transportation</td>
<td>Mainly derived from the empirical setup</td>
<td>Needs to be addressed in order to make the model feasible</td>
</tr>
<tr>
<td>Governing rule issues</td>
<td>Transportation</td>
<td>Derived both empirically and theoretically</td>
<td>Needs to be addressed in order to make the model feasible</td>
</tr>
</tbody>
</table>

### 7. Conclusions

It is concluded that the research effort is to be directed toward the four main areas of transport management, transport planning, planability and optimization. The former issue involves solving the question of the combined effect of the knapsack problems and information quality for the planning and control performance. The latter pertains to the governing rules or design principles that would afford the best system wide performance.

Table 1 summarized the research needs identified in the paper. The combination of the two aspects presented creates a problem. A proper solution will offer a scientific contribution to both theory and practice.

Given the detailed level of feasible planning and its coherence with the idea of FTN, a network optimizing set of rules regarding the distribution of goods between different trucks and network layers ought to be developed. The performance of the final network design and its governing rule should, of course, be evaluated.

The performance resulting system design and governance is to be evaluated with respect to the potential for improvement compared to the state of the art systems of the
present day and the cost (not necessarily a money measure) of achieving an operational FTN.

Acknowledgements

The authors of this paper would like to thank Vinnova for their financial support, Schenker DB Sweden and Volvo Logistics for their cooperation in the collection of empirical data, and Professor Kenth Lumsden for his ideas and support for this paper. Many thanks to all the colleagues from near and far who participated in the seminars and workshops.
References


Hultkrantz, O. (1999) Networks in Road-Based Traffic, Department of Transportation and Logistics, Göteborg, Chalmers University of Technology.


Regulatory instruments to control environmental externalities from the transport sector

Govinda R. Timilsina, Hari B. Dulal

The World Bank
1818 H Street, NW, Washington, DC 20433, USA

Abstract

This study reviews regulatory instruments designed to reduce environmental externalities from the transport sector. We find that the main regulatory instruments used in practice are fuel economy standards, vehicle emission standards and fuel quality standards. While industrialized countries have introduced all three standards with strong enforcement mechanisms, most developing countries have yet to introduce fuel economy standards. The emission standards introduced by many developing countries to control local air pollutants follow either the EU or U.S. standards. Fuel quality standards, particularly for gasoline and diesel, have been introduced in many countries mandating 2 to 10 percents blending of biofuels, 10 to 50 times reduction of sulfur from 1996 levels and banning lead contents. Although inspection and maintenance (I/M) programs are in place in both industrialized and developing countries to enforce regulatory standards, these programs have faced several challenges in developing countries due to a lack of resources. The study also highlights several factors affecting the selection of regulatory instruments, such as countries’ environmental priorities and institutional capacities.

Keywords: Transport sector externalities; Emissions; Regulatory policy instruments.

1. Introduction

Regulatory instruments are legal, enforceable, 'command and control' type instruments aimed at reaching desired, prescribed environmental quality targets or performance standards by regulating the behavior of individuals and/or firms (Seik, 1996). In the transport sector, regulatory instruments induce adjustment of market participants’ behavior (e.g., purchasing more fuel efficient vehicles, lowering operator speeds, optimizing logistics in freight transport, changing the modal split) by establishing suitable incentives (Ahrens, 2008). Examples of these instruments include the following: Corporate Average Fuel Economy (CAFE) standards established in the United States in line with the 1975 Energy Policy Conservation Act; On-Road Vehicle

* Corresponding author: Govinda R. Timilsina (gtimilsina@worldbank.org)
and Engine Emission Regulations established under the 1999 Canadian Environmental Protection Act; and European Union Emission Standards for Light Commercial Vehicles (i.e., Euro 2, Euro 3, and Euro 4 standards). Depending upon the primary objective, existing regulatory instruments target any of the following: (i) direct control of vehicular emissions or exhaust (e.g., emission standards in European Union, the United States, and many developing countries), (ii) reduction of fuel consumption (e.g., CAFE standards in the United States), (iii) cutting vehicle mileage (e.g., authorized mileage rates in the United Kingdom), (iv) lowering traffic congestion (e.g., the odd-and-even license plate rule in Mexico city). Some of these instruments can spur technological innovations. For example, higher CAFE standards can force vehicle manufacturers to produce more fuel-efficient vehicles; emission or exhaust standards mandate vehicles to be fitted with less polluting engines and emission control systems.

The key advantages of regulatory instruments are the directness and relative certainty of outcomes due to compliance measures. They boost economic competitiveness and environmental sustainability (Seik, 1996; Hricko 2004; Bartle and Vass, 2007). Strong regulatory programs and other regulatory efforts have had a significant effect on the control of air pollution in many countries (Ringquist, 1993). Regulatory measures alone, however, might not be sufficient to reduce vehicular emissions to the desired level. Therefore, effective pricing or fiscal policies, sound land use planning and the provision of environmentally sound public transportation systems can reinforce such regulatory measures (Faiz et. al, 1995).

Despite well-established theoretical foundations and wide implementation in the industrialized nations, regulatory policy instruments still present several issues that require further investigation before their widespread introduction in the developing world. The most important issues confronting policy makers in the developing world include, but are not limited to, the following: Which regulatory policy instrument would be the most effective in their context? How to design the implementation mechanisms? Keeping this broad objective in the background, this study presents an in-depth review of various types of regulatory policy instruments, such as fuel economy standards; emissions and exhaust standards; fuel specification standards and inspection and maintenance programs.

The paper is organized as follows: Section 2 briefly introduces various types of regulatory instruments followed by a detailed discussion of fuel economy standards in Section 3. In Section 4, we review vehicle emission standards. Section 5 and Section 6 present, respectively, fuel quality standards and inspection and maintenance programs. Section 7 discusses other laws and regulatory measures to control transport sector emissions. This is followed by discussions on key factors influencing the selection of regulatory instruments in Section 8. Finally, we conclude the paper in Section 9.

2. Types of Regulatory Instruments

Regulatory instruments to control environmental externalities from the transport sector can be classified into different categories using different criteria. For example,
Carbajo and Faiz (1994) classified the instruments into three categories based on the targets of the instruments. These categories are those targeting: (i) vehicle engines (e.g., fuel economy standards, emission standards and inspection and maintenance programs); (ii) fuel quality, such as contents of lead and sulfur and mandatory blending of biofuels; and (iii) transport demand (e.g., traffic management through vehicle bans and designating lanes for high occupancy vehicles). In this paper, we classify the instruments based on the purpose of the instruments. Our classification is as follows: (i) fuel economy standards, which aim at reducing fuel consumption and associated emissions, particularly, CO$_2$; (ii) emission standards which are directly aimed at the reduction of specific emissions released after fuel consumption; (iii) fuel quality standards to reduce or eliminate emission causing elements before the combustion of fuel and (iv) other regulatory measures either discouraging vehicle utilization (e.g., full or partial bans) or encouraging high occupancy of the vehicles (e.g., HOV lanes).

Fuel economy standards refer to standards on vehicle mileage per unit of fuel consumption (i.e., km per liter or miles per gallon). These are common ways to control emissions from the transport sector (Faiz et al., 1995). The CAFE standards introduced in the United States are good examples of fuel economy standards. Fuel economy standards help increase energy efficiency of vehicles, thereby cutting fuel demand and associated emissions. While these standards could be effective in reducing fuel demand and emissions, they do not help in reducing congestion. Fuel economy standards also reduce emissions indirectly by cutting fuel consumption in the supply chain, such as crude oil drilling and production, pipeline and oil refinery. For example, Potter (2003) showed that, in the United Kingdom, out of total emissions from an average car, 76 percent were from fuel usage, 9 percent from manufacturing of the vehicle, and the remaining 15 percent was from losses in the fuel supply system.

Emission standards are aimed at directly reducing emissions, the exhaust coming out of the tail pipes of vehicles. These standards are different from fuel economy standards because they directly control emissions from vehicles, whereas the latter reduce emissions by reducing fuel demand. Fuel economy standards are aimed mainly at reducing fuel consumption and greenhouse gas (GHG) emissions; however, emission standards control local air pollutants, such as suspended particulate matters (SPM), carbon monoxide (CO), volatile organic compounds (VOCs) or non-metallic organic compounds (NMOC), oxides of nitrogen (NOx), etc. While fuel economy standards reduce local air pollution, emission standards do not necessarily reduce fuel consumption as emissions of local air pollutants can be reduced without curtailing fuel consumption by fitting emission controlling devices in vehicles.

Fuel quality standards refer to limits on the content of substances that cause environmental pollution, such as sulfur and lead, in fuel. In order to control emissions of lead and sulfur from vehicular sources, the best approach is to remove these elements from fuels before burning. Regardless of the age or state of repair, lead emissions from all gasoline-fueled vehicles can be eliminated by discontinuing the addition of lead to gasoline. Likewise, emissions of oxides of sulfur (SO$_x$s) can be abated by reducing the sulfur contents of fuels.
3. Fuel Economy Standards

The primary purpose of fuel economy standards is to reduce transport sector fuel demand through vehicle fuel efficiency improvements. A number of countries have introduced fuel economy standards, which help to reduce some types of emissions, such as CO₂, that are directly linked to fuel consumption. While European Union (EU) and Japan have the most stringent fuel economy standards in the world, the United States and Canada had the lowest standards. China has more stringent standards than those of Australia, Canada and the United States (An and Sauer, 2004).

3.1 Corporate Average Fuel Economy Standards in the United States

The CAFE standards require automobile manufacturers to meet stipulated standards for the sales-weighted fuel economy of light duty passenger vehicles sold and to maintain a distinct standard for passenger cars and light trucks (An and Sauer, 2004). Although CAFE is lauded as the main policy instrument to reduce transport sector emissions in the United States, it was, in fact, introduced from an energy security perspective in the mid-1970s. The impetus for CAFE was the oil crisis of 1973 (Proost and Van Dender, 2001). Title V of the Energy Policy and Conservation Act (EPCA), passed by the U.S. Congress in 1975, set automobile fuel efficiency standards for the first time in the United States. CAFE was one of the outcomes of this Act (Faiz et al., 1995; Kirby, 1995).

CAFE standards were initially set for cars and light trucks (light vehicles) (DeCicco, 1995). Currently, vehicles with a gross vehicle weight rating (GVWR) of 1,000 or less are legally obliged to comply with CAFE standards (Komiyama, 2008). Consumers have responded to CAFE standard by switching from large cars to light trucks, a less-regulated class of vehicles (Godek, 1997).

Minimum acceptable standards introduced by the EPCA began in 1978 at 18 mpg for passenger cars. By 1985, the fuel economy standard had increased to 27.5 mpg. Under intense pressure from lobbyists representing auto manufacturers, it was rolled back to 26.5 mpg in 1986. Fuel efficiency standard returned to its previous level of 27.5 mpg in 1989, where it has remained ever since (Kirby, 1995). The United States Congress, in 2007, passed a comprehensive energy bill, The Energy Independence and Security Act of 2007, which includes a provision to achieve fuel economy of 35 miles per gallon (MPG) for new automobiles by 2020 (Komiyama, 2008). The evolution of U.S. CAFE standards illustrates a remarkable improvement in the average on-road fuel economy of new cars and light trucks in the country although the standards do not directly affect vehicles in use. Over time, the U.S. CAFE regulations increased fuel economy from an average 14 mpg in the mid-1970s to 21 mpg in the mid-1990s (Zachariadis, 2006). However, the drag that older vehicles impose on fuel efficiency appears to be quite substantial. The increase in the median age of registered automobiles (5.9 years in 1970 to 7.5 years in 1990 and 9.0 years in 2001), less stringent regulation of light pickup trucks, vans, and sport/utility vehicles has depressed the growth in fuel efficiency

---

2 For some countries/regions (e.g., EU) fuel economy standards are defined in terms of CO₂/GHG emissions per kilometer/miles traveled. Although these standards can be classified as emission standards; we have included them in fuel standards because these standards are implemented through equivalent fuel economy standards.
(Crandall, 1992; de Palma and Kilani, 2008). For example, fuel efficiency of all vehicles on the road has increased by only 34 percent even though the fuel efficiency of new cars increased by 76 percent (Crandall, 1992).

3.2 Fuel Economy Standards in Other Countries

Besides the United States, Australia, Canada, Japan, the European Union, China and South Korea have also specified fuel economy standards for their vehicles.

**Australia:** The Federal Chamber of Automotive Industries (FCAI) first established voluntary fuel economy standards for new passenger cars sold in Australia in 1978 that lasted until 1987. However, those codes failed to achieve the desired targets (CONCAWE, 2006). In 1996, the Ministers for Transport and Primary Industries and Energy endorsed a second voluntary code of practice, which remained in force until July 2001. FCAI members, under the second voluntary code, agreed to reduce the passenger car National Average Fuel Consumption (NAFC) to 8.2 L/100-km (approximately 29 mpg) by the year 2000. In order to maintain the rate of improvement in NAFC achieved for the period up to the year 2000, a third voluntary fuel consumption agreement was reached between the FCAI and the government in 2003, which calls for reduction in fleet average fuel consumption for passenger cars by 18 percent by 2010.

**Canada:** The federal government introduced a voluntary Company Average Fuel Consumption (CAFC) standard in 1976 for the new passenger vehicle fleet. In 1982, the fuel economy standards were made mandatory. These regulations are comparable to the U.S. CAFE standards.

**Japan:** The Japanese government has established a set of fuel economy standards for gasoline and diesel powered light duty passenger and commercial vehicles. The targets to meet the standards are 2005 for diesel and 2010 for gasoline. The standards are based on average vehicle fuel economy by weight class. For gasoline vehicles, it varies from 15 MPG for vehicles weighing more than 2,266 kg to 49.6 MPG for vehicles weighing less than 702 kg (An and Sauer, 2004). By 2010, the average fuel economy of gasoline vehicles is expected to increase by 23 percent from the 1995 level. Regulations for both light duty and heavy-duty diesel vehicles are structured differently. An average regulated emission limit value is used for certification and for production control. This limit is complemented by a slightly higher maximum permissible limit value that must be passed for each vehicle unit (Bauner et al., 2008). Assuming no change in the vehicle mix, the targets for diesel vehicles call for a 14 percent fuel economy improvement compared to the 1995 fleet (11.6 km/l versus 10 km/l).

**European Union (EU):** After an agreement between the European Commission (EC) and the European Automobile Manufacturers Association (ACEA) in 1998 and similar agreements with the Japanese and Korean manufacturers (JAMA and KAMA) in 1999, the EU automobile industry committed to a target by 2008/2009. The major provisions of the ACEA Agreement, signed in March 1998, include a CO₂ emission target of 140 g CO₂/km, representing a 25 percent reduction from the 1995 level of 186 g CO₂/km, to be reached by 2008 with the possibility of an extension of the agreement to 120 g CO₂/km by 2012 (Dieselnet, 2005). The difference between the agreements signed by the European Commission (EC) with the European Automobile Manufacturers Association (ACEA) in 1998 and with the Japanese and Korean manufacturers in 1999 is that the target of 140 g CO₂/km is delayed by one year, to 2009, for, JAMA and KAMA (Dieselnet, 2005).
3.3 Impacts of Fuel Economy Standards on Fuel Consumption and Emissions

The impacts of fuel economy standards on fuel consumption (Geller et al., 1992; Goldberg, 1998; Greene, 1998) and emission reduction (Decicco, 1995) are helpful in assessing the performance of these standards and their suitability for replication in developing countries. Parry et al. (2004) used the Arizona I/M program data collected in 1995 and 2002 on car and truck emissions of volatile organic compounds (VOC), nitrogen oxides (NOx), and carbon monoxide (CO) to study the effects of fuel economy standards on emission rates in the United States. They found emission rates were significantly affected by fuel economy standards in 1995 but not so in 2002. This is mainly because the projected CO, hydrocarbon (HC) and NOx emissions per mile for cars and trucks with certified fuel economy of 20 and 30 mpg are virtually indistinguishable over vehicle lifetimes. Based on their findings, they proposed that lifetime emission rates are equivalent for different cars and for different light trucks. Using a vehicle stock turnover model, Decicco (1995) estimated the effect of enhanced fuel economy standards on gasoline consumption, GHG emissions, and hydrocarbon emissions for light duty vehicles in the United States. The author found that an improvement of 6 percent per year in fuel economy would result in savings of 2.9 million barrel of gasoline per day and 147 million metric tons of annual carbon emission avoidance. Likewise, using in-use emission data collected by remote sensing, Harrington (1997) demonstrated a strong association between better fuel economy and lower emissions of carbon monoxide (CO) and hydrocarbon (HC), which gets even stronger as vehicles age.

Despite the considerable amount of research done on the effects of CAFE on fuel consumption and other related factors, there is no universal consensus on the effects of the CAFE program on the fuel economy of the U.S. vehicle fleet, the overall safety of passenger vehicles, the health of the domestic automobile industry, employment in that industry, and the well-being of consumers (NSC, 2002). Greene (1998) estimated that CAFE standards have led to about a 50 percent increase in on-road fuel economy for light duty vehicles during the period 1975-1995. Improvement in fuel economy forced by the CAFE standards has resulted in an overall decrease in motor fuel expenditure. This means that consumers, in the late 1990s, spent over $50 billion per year less on fuel than what they actually would have spent at 1975 mpg levels. By contributing to increased fuel economy, the CAFE program has reduced dependence on imported oil, improved the nation’s terms of trade, and reduced CO2 emissions relative to what otherwise would have been (NSC, 2002).

Although the overall goal of CAFE regulation has shifted from reducing fuel consumption in a period of high oil prices to reducing harmful emissions, positive environmental gains resulting from CAFE standard has drawn flak from various quarters (Goldberg, 1998). Dowlatabadi et al. (1996) demonstrated that enhanced CAFE standards might have little or no effect on urban air pollution and a less than proportional reduction in GHG emissions. They argued that CAFE is not the most cost effective way of lowering nitric oxide (NO), volatile organic compounds (VOC) and GHG emissions. Portney et al. (2003) asserted that by reducing gallons/mile, the CAFE standards make driving cheaper, which might lead to an overall increase in pollution.

Crandall (1992) ranked the effectiveness of a carbon tax, a petroleum tax, and CAFE standards in terms of their ability to reduce greenhouse gases. He considered a carbon tax to be much more efficient than a petroleum tax. CAFE, according to Crandall,
would cost the economy at least 8.5 times as much as a carbon tax with equivalent
effects on carbon emissions. The inefficiency of the CAFE is mainly because of its
failure to equate the marginal costs of reducing fuel consumption across all uses,
including usage of older vehicles and non-vehicular consumption. Using an empirically
rich simulation model and cost estimates for anticipated fuel economy technologies,
Austin and Dinan (2005), compared the cost of the higher CAFE standards against the
cost of a gasoline tax that would save the same amount of gasoline. Their findings
suggested that a gasoline tax would produce greater immediate savings by encouraging
people to drive less and, eventually, to choose more-fuel-efficient vehicles. Fischer
(2008) and West and Williams (2005) concurred with Austin and Dinan’s assertion that
gasoline taxes are a more efficient means to reduce fuel consumption than mandating
fuel economy increases.

Increased vehicle miles traveled due to enhanced fuel economy is another aspect that
some studies, such as Dowlatabadi et al. (1996), Bamberger (2002) and Portney et al.
(2003), found to be problematic. An increase in VMT also means an increase in
congestion and crash costs (CBO, 2003), and an increase in the overall cost of driving
(Bamberger, 2002). Nivola and Crandall (1995) argued against the effectiveness of
CAFE in reducing vehicle miles traveled and labeled CAFE as a problematic
experiment. They argued that the United States would have saved at least as much oil,
by reducing miles driven in all types and vintages of vehicles, at about a third the
economic cost, if a fee of just 25 cents a gallon had been added to the cost of gasoline
nine years ago. Wang (1994) proposed a marketable permit scheme for light duty
vehicle manufacturers as a more efficient alternative to the existing CAFE standards.
For CAFE to be more effective, Portney et al. (2003) suggested the adoption of tradable
fuel economy (FE) permits among manufacturers, revision of the criterion for
distinction between cars and light trucks, and removal of distinctions between domestic
and imported vehicle fleets.

Several studies (Greene, 1991; NRC 2002; Greene and Hopson, 2003) have measured
the welfare effects of fuel economy regulations by estimating lifetime fuel saving
benefits and subtracting the added vehicle costs from it. Welfare studies widely differ
not only in magnitude but also in the direction of the welfare effect. Kleit (2004)
demonstrated that a long-run MPG increase in the CAFE standard not only causes a
huge welfare loss but that it is also an inefficient instrument for conserving fuel. He
found that a long-run 3.0 MPG increase in the CAFE standard leads to $4 billion of
welfare loss per year and 5.2 billion gallons of gasoline savings per year. He shows that
the same amount of fuel can be conserved with an increase in the gasoline tax of 11
cents per gallon. The overall welfare loss resulting from such an increase would be $290
million per year or about one-fourteenth of the cost imposed in the former case.
Dowlatabadi et al. (1996) argued against further increasing CAFE standards. They
maintain that fuel savings from increasing CAFE are subjected to diminishing returns.
West and Williams (2005) showed that an interaction with the tax-distorted labor
market causes the cost advantage of the gas tax over the CAFE standard to be higher
than anticipated. In such a context, increasing the gas tax would very likely lead to
welfare gain, whereas welfare loss is almost certain if the CAFE standard is tightened.

Table 1 presents the impacts of CAFE standards on fuel savings and job losses. The
CAFE standards might be considered successful in enhancing fuel economy but the
gains achieved through CAFE standards have been undermined by the growth in vehicle
fleet: The policy has not been able to reduce overall fuel demand due to the rapid
growth of the vehicle fleet. Gallagher et al. (2007) pointed out the ineffectiveness of CAFE in terms of ensuring energy security. They argued that, although CAFE standards are politically attractive and induce innovation among other things, it might not be the right policy instrument when it comes to ensuring energy security through reduced fuel consumption. Total motor vehicle fuel consumption in the United States has increased by 60 percent since the enactment of the CAFE program. Enhanced fuel economy standards may have propelled more driving – the so-called “rebound” effect – increasing the total vehicle miles traveled. Greening et al. (2000), however, argued that the increase in travel resulting from the decrease in cost per mile and reduced fuel intensity arising from the CAFE standards is minimal.

Table 1: Macroeconomic and Welfare Impacts of Fuel Economy Standards of US CAFE Standards.

<table>
<thead>
<tr>
<th>Study</th>
<th>Approach</th>
<th>Estimated Impacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dacy et al. (1980)</td>
<td>INFORUM input–output model</td>
<td>A net increase in employment of 140,000 jobs by 1985 due to CAFE standards; job losses in steel, petroleum and gas, and wholesale and retail trade sector are offset by new jobs created in various service industries, plastics, metal stampings, and other sectors.</td>
</tr>
<tr>
<td>Motor Vehicles Manufacturers Association (1990)</td>
<td>Input–output model</td>
<td>The loss of between 159,000 and 315,000 jobs in the motor vehicle industry</td>
</tr>
<tr>
<td>Geller et al. (1992)</td>
<td></td>
<td>Fuel savings of $54 billion (1990 dollar) Increasing the fuel efficiency of passenger cars from 28 mpg in 1990 to 40 mpg in 2000 and 50 mpg in 2010 would create 244,000 by 2010</td>
</tr>
<tr>
<td>Goldberg (1998)</td>
<td></td>
<td>Reduced fuel consumption by 19 million gallons per year; the gasoline tax would have to increase by 780 percent, or 80 cents per gallon, to achieve the same fuel savings as the CAFE standards.</td>
</tr>
</tbody>
</table>


Goldberg (1998) and Parry et al. (2004) argued that welfare gains depend upon myriad factors such as ability of the CAFE to function as a set of internal taxes on fuel inefficient vehicles, subsidies on fuel-efficient vehicles, local pollution, nationwide congestion, traffic accidents, and how consumers value fuel economy technologies and their opportunity costs. CAFE, according to Goldberg (1998), may not fare that badly from a welfare point of view because of its ability to function as a set of internal taxes (on fuel inefficient) and subsidies (on fuel-efficient vehicles) within each firm. Based on the estimates of CAFE’s impact on local pollution, nationwide congestion, and traffic accidents, Parry et al.(2004) found that, contingent upon how consumers value fuel economy technologies and their opportunity costs, higher fuel economy standards can produce anything from significant welfare gains, to very little or no effect, to significant welfare losses. Using marginal oil dependency and carbon externalities value of $0.16 and $0.12 per gallon respectively, they demonstrated that the reduction in fuel demand induced by improved fuel economy is welfare improving only when the marginal external costs of carbon emissions and oil dependency exceed the product of the existing fuel tax and the marginal social value of fuel tax revenues.
4. Vehicle Emission Standards

The implementation of emission standards is the most direct way of reducing emissions per VMT (Walsh, 1992). Without introducing emission standards, policies aimed at reducing fuel consumption and enhancement of fuel economy may not be sufficient to contain local air pollutant from the transport sector (ADB, 2003). Olsson (1994) argued that stringent emission standards lower emissions by forcing the auto industry to derive new vehicle technologies. Emission standards have been introduced in practice in many countries since 1970s. However, levels of emission standards, vehicle coverage, and monitoring and enforcement differ across countries. Here, we briefly discuss a few examples of emission standards introduced in selected countries/states.

4.1 Emission Standards in the United States

In the United States, Congress passed the Clean Air Act in 1970, calling for the first tailpipe emissions standards to control specifically carbon monoxide (CO), volatile organic compounds (VOC), and oxides of nitrogen (NOx). In 1975, the new standards were put into effect with a NOx standard for cars and light duty trucks of 3.1 grams per mile (gpm). In order to make the Act more effective, Congress amended the Act and further tightened emission standards in 1977. The NOx standard, between 1977 and 1979, was reduced from 3.1 gpm to 2.0 gpm for cars. In order to meet the Clean Air Act requirements, the Environmental Protection Agency (EPA) set the first tailpipe standards for light duty trucks at 1.7 gpm in 1979 and for heavier trucks at 2.3 gpm in 1988. Effective in 1988, the standards for light duty trucks were lowered to 1.2 gpm (USEPA, 1999).

Tier 1 Emission Standards in the United States: In 1990, Congress amended the Clean Air Act. Emission standards were further tightened to counter the additional pollution resulting from the increase in vehicle stock. Published as a final rule on June 5, 1991, Tier 1 standards were implemented between 1994 and 1997. Effective in 1994, the NOx standard was set at 0.6 gpm for cars (USEPA, 1999). The Tier 1 vehicle emission standards (0.25 grams per mile non-methane hydrocarbons (NMHC) for light duty vehicles, which were introduced progressively from 1994 onwards in the United States, became obsolete after the 2003 model year with a phase-in implementation of Tier 2 standard schedule from 2004 to 2009 (Gwilliam et al., 2004).

Tier 2 Emission Standards in the United States: The EPA proposed Tier 2 tailpipe emissions standards in 1999 that were to be implemented in 2004. For the first time, both cars and light duty trucks were subject to the same national pollution control system. The same emissions standards apply to all vehicle weight categories. For example, cars, minivans, light-duty trucks, and SUVs have the same emission limit. Tier 2 set the new standard at 0.07 gpm for NOx, a 77 to 86 percent reduction for cars. In order to take full advantage of vehicle emission control technologies, the EPA also proposed a reduction in average sulfur levels to 30 parts per million (ppm) (USEPA, 1999) from the then average of more than 300 ppm. As a comprehensive national control program meant to regulate vehicles and their fuel as a single system, the Tier 2 Emission Standards pursue significant emission reductions (Gwilliam et al., 2004). Tier 2 regulations are more stringent than Tier 1 requirements, and they further extend the
application of the standards to include some of the heavier vehicle categories that were not included in Tier 1 standards (Dieselnet, 2005).

In order to understand how the Tier 2 program works, it is necessary to understand the EPA’s classification of light duty vehicles and trucks. Vehicles and trucks under 8500 lb gross vehicle weight rating (GVWR) are classified as light duty vehicles.

Table 2: Tier 2 Light Duty Full Useful Life Exhaust Emission Standards

<table>
<thead>
<tr>
<th>Bin no</th>
<th>NO_x</th>
<th>NMOG</th>
<th>CO</th>
<th>HCHO</th>
<th>PM</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>0.9</td>
<td>0.28</td>
<td>7.3</td>
<td>0.032</td>
<td>0.12</td>
<td>(1)</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
<td>0.156 (0.230)</td>
<td>4.2 (6.4)</td>
<td>0.018 (0.027)</td>
<td>0.08 (2,3,4)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.3</td>
<td>0.090 (0.180)</td>
<td>4.2</td>
<td>0.018</td>
<td>0.06</td>
<td>(3,5)</td>
</tr>
<tr>
<td>8</td>
<td>0.2</td>
<td>0.125 (0.156)</td>
<td>4.2</td>
<td>0.018</td>
<td>0.02</td>
<td>(2,6)</td>
</tr>
<tr>
<td>7</td>
<td>0.15</td>
<td>0.09</td>
<td>4.2</td>
<td>0.018</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.1</td>
<td>0.09</td>
<td>4.2</td>
<td>0.018</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.07</td>
<td>0.09</td>
<td>4.2</td>
<td>0.018</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.04</td>
<td>0.07</td>
<td>2.1</td>
<td>0.011</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.03</td>
<td>0.055</td>
<td>2.1</td>
<td>0.011</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>0.01</td>
<td>2.1</td>
<td>0.004</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Notes: (1) Bin 11 is only for MDVPs and is available up to and including the model year.  
(2) Bin deleted at the end of 2006 model year (2008 for HLDTs).  
(3) The higher temporary NMOG, CO, and HCHO values apply only to HDLTs and expire after 2008.  
(4) Optional temporary NMOG standard of 0.280 g/mile applies for qualifying LDT4s and MDVPs only.  
(5) Optional Temporary NMOG standard of 0.130 g/mile applies for LDT2s only.  
(6) Higher temporary NMOG standard is deleted at the of 2008 model year.  

Under the Tier 2 programme, manufacturers select a set of full useful life standards from the same row also called (“emission bin” or “bin”) for a given test group of light duty vehicles (LDVs) and light duty trucks (LDTs). The way it works is that, under the “emission bin” approach, manufacturers select a set of emission standards (a bin) to comply with, as a result of which test groups are obliged to meet all standards within that particular bin. For example: If a manufacturer aims for Bin 5 for its light duty diesel vehicles and cannot meet the target, the higher bins in that case allow a safety factor. It is the manufacturer’s responsibility now to offset the higher bin models with similar volumes of lower bin vehicles (CONCAWE, 2006). In addition, the Tier 2 vehicles are obliged to meet the requirements of one of the available “emission bin” and a full life NO_x standard of 0.07 g/miles (CONCAWE, 2006).

California Emission Standards: Among the states in the U.S., California tends to be the leader in imposing increasingly stringent environmental regulations. In 1989, the California Air Resources Board (CARB), in response to severe air pollution problems in Los Angeles and other major cities in California, established stringent, technology-forcing vehicle emission standards to be phased in between the period of 1994 and 2003 (Faiz et. al., 1995). As California began to regulate vehicle emissions earlier than the Federal government, it is treated differently than the other states when it comes to The discretion to adopt its own unique vehicle emissions control program. Under the Clean
Air Act in 1970, California is allowed to set its own emissions standards (ECMT, 2000). The LEV II regulations, which were formally adopted on 5 August 1998 and came into operation on 27 November 1999, are the current standards for California (See table 3 & 4) (CONCAWE, 2006).

Table 3: LEV II Exhaust Emissions Standards-Light and Medium Duty Vehicles.

<table>
<thead>
<tr>
<th>Category</th>
<th>50,000 miles</th>
<th>120,000 miles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NMOG</td>
<td>CO</td>
</tr>
<tr>
<td>LEV</td>
<td>0.075</td>
<td>3.4</td>
</tr>
<tr>
<td>ULEV</td>
<td>0.04</td>
<td>1.7</td>
</tr>
<tr>
<td>SULEV</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: Limits are for intermediate life of 5 yrs or 50,000 or full useful life of 10,000 miles or 10 years. Source: CONCAWE, 2006.

Table 4: LEV II Exhaust Emissions Standards- Medium Duty Vehicles (MDVs).

<table>
<thead>
<tr>
<th>Type (Weight (GVWR), lbs.)</th>
<th>Durability Mileage</th>
<th>Emission category</th>
<th>NMOG</th>
<th>CO</th>
<th>NOx</th>
<th>PM</th>
<th>HCHO</th>
</tr>
</thead>
<tbody>
<tr>
<td>8,500 - 10,000</td>
<td>12,000</td>
<td>LEV</td>
<td>0.195</td>
<td>6.4</td>
<td>0.2</td>
<td>0.12</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ULEV</td>
<td>0.143</td>
<td>6.4</td>
<td>0.2</td>
<td>0.06</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SULEV</td>
<td>0.1</td>
<td>3.2</td>
<td>0.1</td>
<td>0.06</td>
<td>0.008</td>
</tr>
<tr>
<td>10,001 - 14,000</td>
<td>12,000</td>
<td>LEV</td>
<td>0.23</td>
<td>7.3</td>
<td>0.4</td>
<td>0.12</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ULEV</td>
<td>0.167</td>
<td>7.3</td>
<td>0.4</td>
<td>0.06</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SULEV</td>
<td>0.117</td>
<td>3.7</td>
<td>0.2</td>
<td>0.06</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Note: Light duty trucks up to 8,500 lbs GVWR, and medium-duty vehicles that are up to 14,000 lbs GVWR fall under the CA LEV-II standards adopted by California. LEV, ULEV and SULEV stand for, respectively, low-emission vehicles, ultra low-emission vehicles and super ultra-low emission vehicles. The LEV II standards indicate the maximum exhaust emission limits for the intermediate and full useful life of LEVs, ULEVs, and SULEVs. It also includes fuel-flexible, bi-fuel, and duel fuel vehicles when operating on the gaseous or alcohol fuels. Source: CONCAWE, 2006.

4.2 Emission Standards in Canada

The Canadian government, on 12 December 2002, under the Canadian Environmental Protection Act of 1999, published its new On-Road Vehicle and Engine Emission Regulations, which is being applied to vehicles and engines that are manufactured or imported into Canada on or after January 1, 2004. The regulations are similar to
established emission standards and test procedures for on-road vehicles in the United States (CONCAWE, 2006).

4.3 Vehicle Emission Regulations in Europe

In Europe, it was the United Nations Economic Commission for Europe (UN-ECE) that formulated emission regulations in the 1970s and early 1980s (CONCAWE, 2006). The motor vehicle emission regulations developed by the ECE were then adopted by individual member states (Faiz et al., 1995). Although in the early years the European Union (EU) adopted regulations that were almost identical with the ECE equivalents, EU has since become proactive in formulating vehicle emission standards. Under the provisions of the Treaty of Rome, EU member states are legally obliged to follow EU regulations (CONCAWE, 2006). In order to make the existing regulations for light duty vehicles more stringent, the EU council of Ministers, in March 1994, adopted EU Directive 94/12/EC. The new emission limits were applied starting 1 January 1996 for new models and 1 January 1997 for existing models. Unlike previous regulations, it set separate standards for gasoline and diesel-fueled vehicles (CONCAWE, 2006). Tables 5 and 6 below display the EU’s commitment to reducing the transport sector emissions: The EU has, over time, adopted tougher standards for all vehicular pollutants.

Table 5: EU Emission Standards for Passenger Cars (Category M\text{1}\,*), g/km.

<table>
<thead>
<tr>
<th>Tier</th>
<th>Date</th>
<th>CO</th>
<th>HC</th>
<th>HC+NOx</th>
<th>NOx</th>
<th>PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diesel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euro 1†</td>
<td>1992.07</td>
<td>2.72</td>
<td>(3.16)</td>
<td>0.97</td>
<td>(1.13)</td>
<td>0.14</td>
</tr>
<tr>
<td>Euro 2, IDI</td>
<td>1996.01</td>
<td>1</td>
<td>-</td>
<td>0.7</td>
<td>-</td>
<td>0.08</td>
</tr>
<tr>
<td>Euro 2, DI</td>
<td>1996.01*</td>
<td>1</td>
<td>-</td>
<td>0.9</td>
<td>-</td>
<td>0.1</td>
</tr>
<tr>
<td>Euro 3</td>
<td>2000.01</td>
<td>0.64</td>
<td>-</td>
<td>0.56</td>
<td>0.5</td>
<td>0.05</td>
</tr>
<tr>
<td>Euro 4</td>
<td>2005.01</td>
<td>0.5</td>
<td>-</td>
<td>0.3</td>
<td>0.25</td>
<td>0.025</td>
</tr>
<tr>
<td>Euro 5</td>
<td>2009.09*</td>
<td>0.5</td>
<td>-</td>
<td>0.23</td>
<td>0.18</td>
<td>0.005*</td>
</tr>
<tr>
<td>Euro 6</td>
<td>2014.09</td>
<td>0.5</td>
<td>-</td>
<td>0.17</td>
<td>0.08</td>
<td>0.005*</td>
</tr>
<tr>
<td>Petrol (Gasoline)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euro 1†</td>
<td>1992.07</td>
<td>2.72</td>
<td>(3.16)</td>
<td>0.97</td>
<td>(1.13)</td>
<td>-</td>
</tr>
<tr>
<td>Euro 2</td>
<td>1996.01</td>
<td>2.2</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Euro 3</td>
<td>2000.01</td>
<td>2.3</td>
<td>0.2</td>
<td>-</td>
<td>0.15</td>
<td>-</td>
</tr>
<tr>
<td>Euro 4</td>
<td>2005.01</td>
<td>1</td>
<td>0.1</td>
<td>-</td>
<td>0.08</td>
<td>-</td>
</tr>
<tr>
<td>Euro 5</td>
<td>2009.09*</td>
<td>1</td>
<td>0.10</td>
<td>-</td>
<td>0.06</td>
<td>0.005*</td>
</tr>
<tr>
<td>Euro 6</td>
<td>2014.09</td>
<td>1</td>
<td>0.10</td>
<td>-</td>
<td>0.06</td>
<td>0.005*</td>
</tr>
</tbody>
</table>

Notes: * At the Euro 1..4 stages, passenger vehicles > 2,500 kg were type approved as Category N1 vehicles.
† Values in brackets are conformity of production (COP) limits.
\(a\) - until 1999.09.30 (after that date DI engines must meet the IDI limits).
\(b\) - 2011.01 for all models.
\(c\) - and NMHC = 0.068 g/km.
\(d\) - applicable only to vehicles using DI engines.
\(e\) - proposed to be changed to 0.003 g/km using the PMP measurement procedure.
Source: Dieselnet (undated).
### Table 6: EU Emission Standards for Light Commercial Vehicles, g/km.

<table>
<thead>
<tr>
<th>Category</th>
<th>Tier</th>
<th>Date</th>
<th>CO</th>
<th>HC</th>
<th>HC+NOₓ</th>
<th>NOₓ</th>
<th>PM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diesel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N₁, Class I ≤ 1305 kg</td>
<td>Euro 4</td>
<td>2005.01</td>
<td>0.5</td>
<td>-</td>
<td>0.3</td>
<td>0.25</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>Euro 5</td>
<td>2009.09</td>
<td>0.5</td>
<td>-</td>
<td>0.23</td>
<td>0.18</td>
<td>0.005&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Euro 6</td>
<td>2014.09</td>
<td>0.5</td>
<td>-</td>
<td>0.17</td>
<td>0.08</td>
<td>0.005&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>N₁, Class II (1305-1760 kg)</td>
<td>Euro 4</td>
<td>2006.01</td>
<td>0.63</td>
<td>-</td>
<td>0.39</td>
<td>0.33</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Euro 5</td>
<td>2010.09</td>
<td>0.63</td>
<td>-</td>
<td>0.295</td>
<td>0.235</td>
<td>0.005&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Euro 6</td>
<td>2015.09</td>
<td>0.63</td>
<td>-</td>
<td>0.195</td>
<td>0.105</td>
<td>0.005&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>N₁, Class III &gt; 1760 kg</td>
<td>Euro 4</td>
<td>2006.01</td>
<td>0.74</td>
<td>0.46</td>
<td>0.46</td>
<td>0.39</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Euro 5</td>
<td>2010.09</td>
<td>0.74</td>
<td>0.35</td>
<td>0.35</td>
<td>0.28</td>
<td>0.005&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Euro 6</td>
<td>2015.09</td>
<td>0.74</td>
<td>0.215</td>
<td>0.215</td>
<td>0.125</td>
<td>0.005&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Petrol (Gasoline)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N₁, Class I ≤ 1305 kg</td>
<td>Euro 4</td>
<td>2005.01</td>
<td>1</td>
<td>0.1</td>
<td>-</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Euro 5</td>
<td>2009.09</td>
<td>1</td>
<td>0.10&lt;sup&gt;f&lt;/sup&gt;</td>
<td>-</td>
<td>0.06</td>
<td>0.005d&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Euro 6</td>
<td>2014.09</td>
<td>1</td>
<td>0.10&lt;sup&gt;f&lt;/sup&gt;</td>
<td>-</td>
<td>0.06</td>
<td>0.005d&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>N₁, Class II (1305-1760 kg)</td>
<td>Euro 4</td>
<td>2006.01</td>
<td>1.81</td>
<td>0.13</td>
<td>-</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Euro 5</td>
<td>2010.09</td>
<td>1.81</td>
<td>0.13&lt;sup&gt;g&lt;/sup&gt;</td>
<td>-</td>
<td>0.075</td>
<td>0.005d&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Euro 6</td>
<td>2015.09</td>
<td>1.81</td>
<td>0.13&lt;sup&gt;g&lt;/sup&gt;</td>
<td>-</td>
<td>0.075</td>
<td>0.005d&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>N₁, Class III &gt; 1760 kg</td>
<td>Euro 4</td>
<td>2006.01</td>
<td>2.27</td>
<td>0.16</td>
<td>-</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Euro 5</td>
<td>2010.09</td>
<td>2.27</td>
<td>0.16&lt;sup&gt;h&lt;/sup&gt;</td>
<td>-</td>
<td>0.082</td>
<td>0.005d&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Euro 6</td>
<td>2015.09</td>
<td>2.27</td>
<td>0.16&lt;sup&gt;h&lt;/sup&gt;</td>
<td>-</td>
<td>0.082</td>
<td>0.005d&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Notes: † For Euro 1/2 the Category N₁ reference mass classes were Class I ≤ 1250 kg. Class II 1250-1700 kg, Class III > 1700 kg.

- a - until 1999.09.30 (after that date DI engines must meet the IDI limits).
- b - 2011.01 for all models.
- c - 2012.01 for all models.
- d - applicable only to vehicles using DI engines.
- e - proposed to be changed to 0.003 g/km using the PMP measurement procedure.
- f - and NMHC = 0.068 g/km.
- g - and NMHC = 0.090 g/km.
- h - and NMHC = 0.108 g/km.


The European emissions standards have become stricter with the adoption of newer Euro limit values. The EU has gradually tightened catalyst-forcing standards for new gasoline-fueled cars (also called Euro 1 standards) since its adoption in the early 1990s and adopted Euro 2, Euro 3, and Euro 4 in 1996, 2000, and 2005 respectively. It also adopted similar requirements for diesel cars and light and heavy commercial vehicles (ADB, 2003). In response to the ongoing planned and probable control measures across
the European Union (EU), by the year 2010, vehicular emission in Europe are expected to fall markedly (Reis et al., 2000). The maximum permissible limits set by Euro 3 called for 30 percent reduction of carbon monoxide (CO), hydrocarbons (HC) and nitrogen oxides (NOx) and 80 percent reduction of particulate matter (PM) emissions. Euro 5 regulations, which new models were obliged to meet starting October 1, 2008, and new registrations of vehicle models certified earlier are supposed to meet starting October 1, 2009, are even more stringent. NOx emission limits are further reduced, by 60 percent compared to Euro 3 (Bauner et al., 2008). Because of the voluntary agreement between the European Automobile Manufacturers Association (ACEA) and the European Commission, the former are obliged to reduce the fuel consumption and average unit emissions of CO\textsubscript{2} of new private cars, both gasoline and diesel, by 21 percent from the period of 1995 to 2008 (Joumard, 2005).

Emission standards alone will not be able to constrain car usage and associated emissions. With an increase in living standards, consumer preferences do shift considerably. In the EU, while Gross Domestic Product (GDP) grew 2.5 percent between 1970 and 1997, annual passenger and freight transport averages increased by an average of 2.8 and 2.6 percent (Walsh, 2000). A gradual shift in consumers’ preference towards new low emission car purchases might be able to slow down the rise in emissions level but more cars on roads also means more congestion and emissions. In addition to the enforcement of stringent emission standard, the following measures should be implemented to improve the effectiveness of emissions control policies: (i) measures such as the use of renewable or non-fossil based fuels and alternative technologies such as fuel cells and gasoline-electric hybrid engine; (ii) shift to less energy intensive modes and reductions in travel, (iii) technological improvements in fuel economy; and, (iv), an increase in load factors (Scholl et. al., 1996; Dargay and Gately, 1997; Kosugia et. al., 2005).

4.4 Vehicle Emissions Standards in Latin America

Like other developing countries, Latin American countries have witnessed rapid growth in transport sector emissions. Urban air quality has deteriorated with an increase in the number of vehicles on urban roadways. In Buenos Aires, for example, the transport sector accounts for over 99 percent of CO emissions and 46 percent of the NOx emissions (Venegas and Mazzeo, 2006). The situation in Brazil is quite similar. In 2004, transport sector emissions accounted for 46 percent of total HC and 98 percent of total CO in the Saõ Paulo Metropolitan Area (SPMA) (Vivancoa and Andradeb, 2006). In Santiago, Chile, older cars and diesel-powered vehicles are the main contributors to CO and NOx concentrations. Between 1990 and 2000, they accounted for 65 percent of total urban air emissions (Jorquera, 2002). In Mexico City, the transport sector accounted for 98 percent of total CO emissions, 40 percent of total HC emissions, and 81 percent of total NOx emissions (Molina and Molina, 2002).

In response to rapidly deteriorating urban air quality, Latin American countries have initiated or adopted emission standards. The stringency of the standards, however, varies across countries/cities depending upon the level of air pollution and other factors. As outlined in Table 7, many Latin American countries have imposed complete or partial bans on used vehicles imports. Despite a huge market for used vehicles, countries such as Argentina, Brazil, Chile, Colombia, Ecuador, Mexico, Paraguay,
Uruguay, and Venezuela have completely banned used vehicle imports (Pelletiere and Reinert, 2002).

Table 7: Latin America Vehicle Standards.

<table>
<thead>
<tr>
<th>Country</th>
<th>Imported</th>
<th>Locally Manufactured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>Only new vehicles, equipped with emission control technologies according to Euro 3 standards</td>
<td>As of 2006, new light duty vehicles must comply with Euro 3, Euro 4 as of 2009, likewise for new diesel trucks and buses.</td>
</tr>
<tr>
<td>Brazil</td>
<td>No importation of used vehicles; imported new vehicles must meet Euro 4 standards</td>
<td>Vehicle emissions standards set by IBAMA, based on Euro standards: Euro 2 implemented in 1993, Euro 4 planned for 2008 equivalent to PROCONVE IV standard, and Euro 4 in 2009. All new trucks and buses must be Euro 4 in 2009.</td>
</tr>
<tr>
<td>Colombia</td>
<td>Importation of used vehicles is banned.</td>
<td>Light duty petrol vehicles must meet USEPA 1987 standards. New vehicles must comply with Euro 1; heavy duty diesel vehicles must comply with equivalent of USEPA 1994 standards for buses and 1991 standards for other vehicles. New buses must comply with Euro 2, other new heavy duty vehicles with Euro 1.</td>
</tr>
<tr>
<td>Ecuador</td>
<td>Importation of used vehicles is banned. Model 2000 and newer cars must possess catalytic converters</td>
<td>New light duty petrol vehicles must meet USEPA 1987 standards or Euro 1; new heavy duty diesel vehicles must comply with USEPA 1994 standards or Euro 2.</td>
</tr>
<tr>
<td>Mexico</td>
<td>Vehicle maximum 10 years, must have a gasoline engine, and must be equipped with a catalytic converter</td>
<td>Since 1993, heavy duty diesel vehicles must meet one of these standards: US 1998, US 2004, Euro 3, or Euro 4. All light duty and passenger vehicles must meet US Tier 1, except on NOx (levels vary) and PM (applies only to diesel).</td>
</tr>
<tr>
<td>Paraguay</td>
<td>Importation of used vehicles is banned.</td>
<td></td>
</tr>
<tr>
<td>Venezuela</td>
<td>Importation of used vehicles is banned.</td>
<td>Emissions testing in certain areas, with fines for violators.</td>
</tr>
</tbody>
</table>


Table 8 shows the emission standards adopted by selected countries in Latin America. Argentina, Brazil, and Chile have chosen to adopt EU standards, whereas Colombia, Ecuador, and Mexico have provided flexibility by adopting both the U.S. equivalents and EU standards. As compared to Argentina and Brazil, Chile, and Mexico have introduced more stringent emission standard.
### Table 8: Emission Standards in selected Latin American countries.

<table>
<thead>
<tr>
<th>Country</th>
<th>Vehicle type</th>
<th>Effective Date</th>
<th>CO (g/km)</th>
<th>HC (g/Km)</th>
<th>NOx (g/km)</th>
<th>PM (g/km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>New Vehicles</td>
<td>1/1/1995</td>
<td>12</td>
<td>1.2</td>
<td>1.4</td>
<td>0.373</td>
</tr>
<tr>
<td></td>
<td>All imports</td>
<td>1/1/1997(2)</td>
<td>2</td>
<td>0.3</td>
<td>0.6</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>All new regular</td>
<td>1/1/1998</td>
<td>6.2</td>
<td>0.5</td>
<td>1.43</td>
<td>0.16(3)</td>
</tr>
<tr>
<td>Brazil</td>
<td>Cars</td>
<td>1/1/1992</td>
<td>24</td>
<td>2.1</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Light Duty</td>
<td>01/01/1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>Passenger cars</td>
<td>1/1/1995</td>
<td>2.11</td>
<td>0.25</td>
<td>0.62</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>Light &amp; Medium Duty (gvw &lt; 3860 kg)</td>
<td>1/2/1995</td>
<td>6.2</td>
<td>0.5</td>
<td>1.43</td>
<td>0.16</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>Gasoline passenger cars and light duty vehicles</td>
<td>1/1/1995</td>
<td>5.7</td>
<td>0.25</td>
<td>0.63</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>&lt; 1800 kg</td>
<td>1/1/1995</td>
<td>6.2</td>
<td>0.5</td>
<td>1.1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1800-2800 kg</td>
<td>1/1/1995</td>
<td>19.2</td>
<td>1.2</td>
<td>10.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>&gt;6400 kg</td>
<td>1/1/1995</td>
<td>49.8</td>
<td>2.3</td>
<td>10.6</td>
<td>-</td>
</tr>
<tr>
<td>Mexico</td>
<td>Cars</td>
<td>1/1/1993</td>
<td>3.4</td>
<td>0.41</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>light duty vehicles</td>
<td>1/1/1994</td>
<td>14</td>
<td>1</td>
<td>2.3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>gvw &lt; 6012 lb</td>
<td>1/1/1994</td>
<td>14</td>
<td>1</td>
<td>2.3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>gvw 6013-6614 lb</td>
<td>1/1/1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: (1) Diesel Vehicles only; (2) 1/1/99 for all new registrations; (3) PM 0.31 g/km for vehicles < 1700kg
Source: CONCAWE (2006)

The introduction of emission standards for both new and old cars, along with travel demand management programs, and regulatory measures such as vehicle inspection and maintenance programs (I/M), fuel specification, etc., have reduced vehicular emissions in Latin American countries. For example, in Mexico City, the total daily CO and NOx emissions from light and medium gasoline vehicle in 2000, were 48 percent and 26 percent lower, respectively, from 1998 levels (Schifter et. al., 2005).

### 4.5 Vehicle Emissions Standards in Asia

Emission standards have been widely implemented in Asia. Some Asian countries (e.g., Singapore, Hong Kong) have introduced and strictly enforced stringent emission standards (Seik, 1996); others are yet to get there. Besides lower standards, strict enforcement is a major challenge in Asia. For example, China’s current limit (Euro II), as compared to the United States, is 26 percent higher for carbon monoxide and double for hydrocarbons. However, the proposed Euro II standards have not been met due to weak enforcement (Zhao, 2004).
Table 9: Exhaust Emission Regulations in Selected Countries.

<table>
<thead>
<tr>
<th>Country</th>
<th>Vehicle type</th>
<th>Fuel</th>
<th>Effective date</th>
<th>Equivalent Emission Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh</td>
<td>Light &amp; Heavy Duty Light &amp; Heavy Duty</td>
<td>Gasoline &amp; Diesel</td>
<td>2006, 2006</td>
<td>Euro II, Euro I</td>
</tr>
<tr>
<td>China$^{(1)}$</td>
<td>Light Duty ($&lt;$3.5t)$^{(2)}$, National</td>
<td>Gasoline &amp; Diesel</td>
<td>1993</td>
<td>ECE 15.03 with higher limits</td>
</tr>
<tr>
<td></td>
<td>Passenger Cars &amp; Light Duty (Beijing &amp; Shanghai)</td>
<td></td>
<td>July, 1999</td>
<td>Euro I</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>Light Duty$^{(3)}$</td>
<td>Gasoline &amp; Diesel</td>
<td>01/01/2006</td>
<td>Euro IV</td>
</tr>
<tr>
<td>India</td>
<td>Light Duty- National Light Duty-Delhi region</td>
<td>Gasoline &amp; Diesel</td>
<td>2000, 04/00</td>
<td>91/441/EEC$^{(4)}$</td>
</tr>
<tr>
<td>Indonesia</td>
<td>Gasoline engines Diesel engines</td>
<td>Gasoline &amp; Diesel</td>
<td>2005, 2005</td>
<td>Euro II</td>
</tr>
<tr>
<td>Malaysia</td>
<td>Light Duty Gasoline Diesel</td>
<td>01/01/00, 01/01/00</td>
<td>94/12/EEC</td>
<td>94/12/EEC</td>
</tr>
<tr>
<td>Nepal</td>
<td>Light Duty-Imported</td>
<td>Gasoline Diesel</td>
<td>01/02</td>
<td>Euro I</td>
</tr>
<tr>
<td>Philippines</td>
<td>Light Duty Medium &amp; heavy duty</td>
<td>Gasoline Diesel</td>
<td>01/01/97, 01/01/97</td>
<td>ECE R 15-04$^{(5)}$, ECE R 49-01</td>
</tr>
<tr>
<td>Singapore</td>
<td>Light Duty Gasoline Diesel</td>
<td>01/2001, 10/2006</td>
<td>Euro II</td>
<td>Euro IV</td>
</tr>
<tr>
<td>South Korea</td>
<td>Gasoline Diesel</td>
<td></td>
<td>01/01/2003</td>
<td>US procedures ECE R 49</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>Gasoline Diesel</td>
<td></td>
<td>01/01/2003</td>
<td>Euro II</td>
</tr>
<tr>
<td>Taiwan</td>
<td>Passenger Cars$^{(6)}$, Light duty$^{(6)}$</td>
<td>Gasoline Diesel</td>
<td>07/90</td>
<td>US 1984 Limits US 1984 LDT</td>
</tr>
<tr>
<td>Thailand</td>
<td>Light Duty All$^{(7)}$</td>
<td></td>
<td>25/08/2001</td>
<td>96/69/EC</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td></td>
<td></td>
<td>25/08/2001</td>
<td>ECE R 15.03 equivalent</td>
</tr>
</tbody>
</table>

Notes:  
$^{(2)}$A government notice, posted on 27 June 2001, required the immediate cessation of production of carbureted vehicles. Production was halted immediately and sales were banned from 1 September 2001.  
$^{(3)}$Euro 3 or equivalent standards will apply to certain class of vehicles under 3.5 tones on or after 1 January 2002. From 1 January 2006, LD diesel must comply with California regulations. Euro 4 introduced from 01/01/2006 for vehicles up to 2.5 tones, extending to 3.5 tones from 01/01/2007.  
$^{(4)}$Employs a modified Indian Driving cycle similar to the ECE15+EUDC cycle, except that the maximum speed is limited to 90 km/h.  
$^{(5)}$Evaporative emission for spark ignition engines shall not exceed 2.0 grams per test. Crankcase emissions should be eliminated.  
$^{(6)}$Evaporative emission for spark ignition engines shall not exceed 2.0 grams per test.  
$^{(7)}$Proposed to the National Environment Board for implementation as follows: RM $\leq$1305 kg from January 2003; RM $>1305$ Kg from 1 January 2004. Implementation of Row B of 98/69/EC (Euro 4) is under discussion.  
Source: CONCAWE (2006)

Table 9 illustrates exhaust emissions regulations in selected Asian countries. Countries such as Bangladesh, India, Indonesia, Sri Lanka, Nepal, and Singapore have introduced Euro standards, whereas Malaysia, Philippines, South Korea, Taiwan, and Saudi Arabia have implemented U.S. emissions regulations.
In order to combat deteriorating urban air quality, China has adopted aggressive vehicle emissions standards. It imposed emissions standards equivalent to Euro 1 in 2000 and aims at meeting current European emissions standards, with a lag of about 4–6 years. (Bauner et al., 2008). The existing vehicle emissions standards adopted in Beijing are similar to Euro 2 standards (Deng, 2006). Euro 4 standards will kick in starting 2010 (Liu et al., 2008).

Like mainland China, Taiwan, too, has taken some bold steps towards containing transport sector emissions. The first stage emission standards for gasoline cars were introduced on 1 July 1987. In Taiwan, all passenger cars must pass emission standard tests for CO and HC during the idle phase at 0.5% and 100 ppm, respectively, for new cars and 1.2% and 220 ppm for in-use cars. Vehicle regulation requires all new passenger vehicles to have exhaust catalyst. It also requires all vehicles to undergo annual I/M tests to pass the emission standards (Chiang et al., 2008).

Japan is another Asian country that has taken strong measures towards vehicular emission control. In addition to various fiscal instruments, Japan has put in place tough regulatory standards. Its emission standards are clearly on par with standards adopted in Europe and the United States. There are two sets of standards: the first one aimed at reducing pollution from vehicles below 1250 Kg and the second for vehicles weighing more than 1250 Kg. Table 10 illustrates the differences in these two sets of standards. Japan’s Central Environmental Council (CEC) published its third report on “Future policy for motor vehicle exhaust emission reduction” in December of 1998. It called for a further strengthening of NO\textsubscript{x} and PM limits for diesel engines in two stages and led to 25-30 percent reduction in NO\textsubscript{x} emission and 28-35 percent reduction in PM emission from 2002-2004, depending on vehicle category. It required 70 percent reduction in HC and CO emissions (CONCAWE, 2006).

**Table 10: Japanese Emission Standards for Diesel Passenger Cars, g/km.**

<table>
<thead>
<tr>
<th>Vehicle Weight</th>
<th>Date</th>
<th>Test</th>
<th>CO</th>
<th>HC</th>
<th>NO\textsubscript{x}</th>
<th>PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1250 kg*</td>
<td>2005\textsuperscript{a}</td>
<td>JC08\textsuperscript{c}</td>
<td>0.63</td>
<td>0.024\textsuperscript{d}</td>
<td>0.14</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td></td>
<td>0.63</td>
<td>0.024\textsuperscript{d}</td>
<td>0.08</td>
<td>0.005</td>
</tr>
<tr>
<td>&gt; 1250 kg*</td>
<td>2002\textsuperscript{a}</td>
<td></td>
<td>0.63</td>
<td>0.12</td>
<td>0.3</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>2005\textsuperscript{b}</td>
<td>JC08\textsuperscript{c}</td>
<td>0.63</td>
<td>0.024\textsuperscript{d}</td>
<td>0.15</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td></td>
<td>0.63</td>
<td>0.024\textsuperscript{d}</td>
<td>0.08</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Notes: \* - equivalent inertia weight (EIW); vehicle weight of 1265 kg.
\textsuperscript{a} - 2002.10 for domestic cars, 2004.09 for imports; \textsuperscript{b} - full implementation by the end of 2005
\textsuperscript{c} - full phase-in by 2011; \textsuperscript{d} - non-methane hydrocarbons.

5. Fuel Quality Standards

Fuel quality standards play a crucial role in protecting public health and the environment from transport sector emissions. It is often viewed as an important component of an overall plan to improve air quality. Cleaner fuels have an immediate impact on both new and existing vehicle fleets. There is a close relationship between
fuel quality and emission control technologies, and it is also important for the successful adoption of stringent vehicle emission standards. The reduction of sulfur to near-zero levels is prerequisite for any air pollution reduction strategy to bear fruits (Hao et al., 2006; Blumberg et al., 2003).

Realizing the importance of cleaner fuel, countries started reducing the level of lead and sulfur in fuel in early the 1990s. Starting January 1995, leaded gasoline sales were banned in the United States. The maximum amount of lead permitted in unleaded gasoline in the United States is 0.013 grams/liter (CONCAWE, 2006). The Alliance of Auto Manufacturers, which represents the auto industry, supported a gasoline sulfur control program in 2004 and agreed to reduce sulfur content to “near-zero” levels (less than 5 mg/kg) by 2007 (CONCAWE, 2006). Similarly, leaded gasoline was banned in the EU effective from 1 January 2000, although some countries like Greece, Italy, and Spain had to be granted a grace period (Gwilliam et al., 2004). EU Directive 2003/17/EC introduced a new sulfur requirement for both gasoline and diesel (maximum 10 mg/kg) and called for the complete penetration of these fuels from 1 January 2009 (CONCAWE, 2006).

Table 11: Gasoline Specification-Selected Developing Countries.

<table>
<thead>
<tr>
<th>Country</th>
<th>Property</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RON (Value Min.)</td>
<td>Sulfur (mg/kg or ppm, Max)</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>80</td>
<td>95</td>
</tr>
<tr>
<td>India</td>
<td>88</td>
<td>93</td>
</tr>
<tr>
<td>Malaysia</td>
<td>92</td>
<td>97</td>
</tr>
<tr>
<td>Philippines</td>
<td>-</td>
<td>93</td>
</tr>
<tr>
<td>Pakistan</td>
<td>80</td>
<td>87</td>
</tr>
<tr>
<td>Thailand</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Kenya</td>
<td>83</td>
<td>93</td>
</tr>
<tr>
<td>Tanzania</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Argentina</td>
<td>83</td>
<td>-</td>
</tr>
<tr>
<td>Bolivia</td>
<td>85</td>
<td>95</td>
</tr>
<tr>
<td>Colombia</td>
<td>81</td>
<td>87</td>
</tr>
<tr>
<td>El Salvador</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Guatemala</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Honduras</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Mexico</td>
<td>-</td>
<td>95</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Panama</td>
<td>87</td>
<td>91</td>
</tr>
<tr>
<td>Paraguay</td>
<td>85</td>
<td>97</td>
</tr>
</tbody>
</table>


Table 11 illustrates specifications for unleaded gasoline in selected developing countries. Fuel quality regulations and specifications vary from one country to another. In countries like Mexico, the maximum allowable limit of sulfur in fuel is far lower than in countries such as Pakistan, India, Guatemala, El Salvador, Honduras, Malaysia, and Tanzania. In sub-Saharan Africa, lead was banned on 1 January 2006; the maximum
allowable limit is 13 mg/l (CONCAWE, 2006). Sulfur limits, especially in diesel, tend to be very high in Pakistan, Malaysia, India, Bangladesh, Thailand, El Salvador, Guatemala, Honduras, and Nicaragua.

China is taking aggressive steps towards containing hazardous components in fuel. By 1998, the local government in Beijing successfully phased out leaded gasoline. At present, sulfur content ranges from 300 ppm to 500 ppm for gasoline and from 500 ppm to 800 ppm for diesel fuel in Beijing (Hao et. al., 2006). Since eliminating lead as an octane booster in gasoline is a relatively low cost measure with high returns in terms of public health, Gwalliam et al. (2004) suggested that it should be a high priority for all countries that have not yet eliminated lead from gasoline.

The emission of sulfur dioxide from diesel used in heavy vehicles is one of the main environmental concerns in most countries around the world. Hence, these countries have imposed standards on the sulfur content of diesel. Table 12 presents existing or planned standards for the sulfur content of diesel in selected countries. As can be seen from the table, sulfur standards for diesel have been rapidly stiffened in many countries over the last decade. For example, the standards in the United States, Japan and European Union have been reduced to 50 ppm in 2005 from 500 ppm in 1996. In Australia, the standards have been reduced to 50 ppm in 2006 from 2000 ppm in 1996. The standards stiffened further to 10 ppm in Japan and European Union. In some developing countries, such as, India, Philippines, Vietnam, the standards for diesel sulfur content were reduced by 10 times during the 1996-2005 period.

Table 12: Existing and Planned Standards for Diesel Sulfur Contents in Selected Countries.
Unit: PPM (milligram of sulfur per kilogram of diesel).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>EU</td>
<td>500</td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Japan</td>
<td>500</td>
<td>50</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>2,000</td>
<td>500</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangladesh</td>
<td>&gt;5000</td>
<td>5,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambodia</td>
<td>&gt;5000</td>
<td>2,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>5,000</td>
<td>2,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>5,000</td>
<td>2,500</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>350</td>
</tr>
<tr>
<td>Indonesia</td>
<td>5,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Korea</td>
<td>2,000</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td>5,000</td>
<td>3,000</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pakistan</td>
<td>10,000</td>
<td>5,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>5,000</td>
<td>2,000</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>5,000</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>5,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3,000</td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>2,500</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td>10,000</td>
<td>2,000</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Krylov et al. (2005).
Although the costs and benefits associated with sulfur reduction vary from region to region, depending on the state of existing refineries, fuel quality, and emissions standards, the cost of sulfur reduction is affordable (Blumberg et. al., 2003). Some countries that import petroleum products might find it hard to maintain the required quality due to the lack of their own refineries. Consequently, developing countries without their own refineries may not be in a position to enforce fuel standard related regulations. Nepal, for example, lacking its own refinery, is dependent on imported petroleum products and is experiencing severe air pollution problems related to the high levels of benzene in imported gasoline (Kiuru, 2002).

Table 13: Biofuels Blending Mandates.

<table>
<thead>
<tr>
<th>Country</th>
<th>Ethanol</th>
<th>Biodiesel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>E2 in New South Wales, increasing to E10 by 2011; E5 in Queensland by 2010</td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>E5 by 2010</td>
<td>B5 by 2010</td>
</tr>
<tr>
<td>Bolivia</td>
<td></td>
<td>B2.5 by 2007 and B20 by 2015</td>
</tr>
<tr>
<td>Brazil</td>
<td>E22-E25</td>
<td>B2 by 2008 and B5 by 2013</td>
</tr>
<tr>
<td>Canada</td>
<td>E5 by 2010; E7.5 in Saskatchewan and Manitoba; E5 by 2007 in Ontario</td>
<td>B2 by 2012</td>
</tr>
<tr>
<td>China</td>
<td>E10 in 9 provinces</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>E10</td>
<td>B5 by 2008</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>E15 by 2015</td>
<td>B2 by 2015</td>
</tr>
<tr>
<td>Germany</td>
<td>E2 by 2007</td>
<td>B4.4 by 2007; B5.75 by 2010</td>
</tr>
<tr>
<td>India</td>
<td>E10 in 13 states/territories</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>E1</td>
<td>B1</td>
</tr>
<tr>
<td>Malaysia</td>
<td></td>
<td>B5 by 2008</td>
</tr>
<tr>
<td>Paraguay</td>
<td></td>
<td>B1 by 2007, B3 by 2008, and B5 by 2009</td>
</tr>
<tr>
<td>Peru</td>
<td>E7.8 by 2010 nationally; starting regionally by 2006</td>
<td>B5 by 2010 nationally; starting regionally by 2008</td>
</tr>
<tr>
<td>Philippines</td>
<td>E5 by 2008; E10 by 2011</td>
<td>B1 by 2008; B2 by 2011</td>
</tr>
<tr>
<td>South Africa</td>
<td>E8-E10 (proposed)</td>
<td>B2-B5 (proposed)</td>
</tr>
<tr>
<td>Thailand</td>
<td>E10 by 2007</td>
<td>3 percent share by 2011</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>E2.5 by 2008; E5 by 2010</td>
<td>B2.5 by 2008; B5 by 2010</td>
</tr>
<tr>
<td>United States</td>
<td>E10 in Iowa, Hawaii, Missouri, and Montana; E20 in Minnesota; E2 in Louisiana and Washington State</td>
<td>B5 in New Mexico; B2 in Louisiana and Washington State</td>
</tr>
<tr>
<td>Uruguay</td>
<td>E5 by 2014</td>
<td>B2 (2008-2011) and B5 by 2012</td>
</tr>
</tbody>
</table>

Note: Targets with no dates are already in place except in some U.S. states where the targets are expected to be effective in future years. There are other countries with future indicative targets that are not shown here.
Another important standard imposed on fuels in many countries is the minimum blending requirement of gasoline and diesel with ethanol and bio-diesel, respectively. Although energy security could be the primary purpose of such blending, reducing environmental externalities, particularly CO\(_2\) emissions, is an equally important benefit. Table 13 presents examples of biofuels blending regulations in selected industrialized and developing countries. Most of these regulations were enacted quite recently, and they typically call for the blending of 10–15 percent ethanol with gasoline or the blending of 2–5 percent biodiesel with diesel. The provinces of British Columbia and Quebec in Canada have also announced that they would mandate ethanol blending but exact blending percentages are yet to be stipulated. Brazil has mandated the blending of biofuels for 30 years through its “ProAlcool” program; while the blending shares for ethanol were adjusted occasionally, they have remained in the 20-25 percent range.

6. Vehicle Inspection and Maintenance Programs

Inspection and maintenance (I/M) programs are largely devised to identify primary “gross polluters” and ensure that they are retrofitted or retired. Be it developed or developing countries, vehicles that are not properly maintained are responsible for a large fraction of total transport sector emissions. Based on a cross country study of CO and HC emissions from over 200,000 vehicles in the USA, Canada, Mexico, the UK, and Sweden, Guenther et al. (1994) found that less than 10 percent of the fleet, which are referred to as “gross polluters,” are responsible for half of the total emissions. Likewise, around 10–12 percent of the existing vehicle fleet accounted for about 50 percent of transport sector CO emissions in Nepal from 2001-2002 (Faiz et al., 2006). Therefore, the problem of a small percentage of ill-maintained vehicles diluting the gains made through higher fuel, emissions, and fuel economy standards is not a developed or developing countries’ problem; it is a global problem that calls for innovative ways to discourage “gross polluters” from getting on the roadways.

Although I/M programs have been widely implemented in both the developed and developing word, there is no universal consensus on the use of I/M programs to regulate vehicle emissions. Faiz et al. (1990) and Mage and Walsh (1992) emphasized the importance of I/M programs. According to Faiz et al. (1990), without a rigorous I/M program, smoke and particulate emissions from often overloaded and poorly maintained diesel-powered vehicles cannot be controlled in developing countries. Mage and Walsh (1992) argued that I/M programs are critical for controlling emissions from both new car and in-use vehicles. Gwalliam (2004) and Kebin and Chang (1999), based on experiences from Mexico City and China, considered I/M programs a success. The I/M system introduced in Mexico city with high volume, centralized test centers is an example of a successful program on a large scale (Gwalliam, 2004). In Beijing, according to Kebin and Chang (1999), emissions decreased a total of 28 to 40 percent, and in Shanghai, CO and HC emission concentrations decreased on average by 39 percent. Like in Beijing, the I/M program introduced in 1992 in the Lower Fraser Valley of the Canadian province of British Columbia, led to reduction in HC emissions by 20 percent, CO by 20 percent, and NO\(_x\) by 1 percent (Faiz et al., 1995). Contrary to the aforementioned studies, Hubbard (1997) argued that the existing I/M programs in the United States have generated, at most, small environmental benefits.
Despite some criticisms, I/M programs have been widely implemented. In the United States, California was the first state to implement a wide-ranging test and repair I/M program in 1984. It required gasoline-powered automobiles to pass inspections every two years (Faiz et al., 1995). In other states, depending upon the state’s performance standards, motorists have to satisfy I/M requirements (Harrington et al., 2000).

Within the European Union, the member states have implemented the requirements of the Roadworthiness Framework Directive. It requires vehicle owners to go for a compulsory vehicle inspection and is enforced to ensure the necessary maintenance and upkeep of vehicles (CCAP, 2004). EU Directive 96/96/EC regulates I/M programs and safety inspections. The directive also provides some leeway to the member states in terms of: (i) setting higher frequency of tests; (ii) making the testing of optional equipment compulsory; (iii) expanding test requirements to other classes of vehicles; and, (iv) prescribing additional or more stringent tests (USAID, 2004).

In Australia, a pilot I/M scheme was introduced in July 1998 in the greater Sydney area. Its main aim was to include all light duty vehicle by the year 2000. The main goal of the National In-service Emissions (NISE2) study was to establish a primary phase and a main phase testing that would aid in the establishment of the current emissions performance of light duty petrol vehicles (CONCAWE, 2006). The primary phase was designed to develop and validate reliable emission tests for light duty gasoline vehicles that are based on “real world” driving patterns. It was intended to provide the basic tools for use in the main phase for generating a more accurate and representative measure of the actual amount of pollutants emitted from the light duty gasoline fleet (CONCAWE, 2006).

China’s I/M programs require regular inspections, which include yearly inspections, first-class maintenance, second class maintenance, and vehicle overhaul. In big cities such as Beijing, Shanghai, and Guangzhou, I/M programs have been effective, to a large degree, in lowering vehicle emissions (Kebin and Chang, 1999).

Despite their emission reduction potential, I/M programs have certain limitations, which are primarily on two fronts: (i) inefficient use of resources and inconvenience to motorists; and, (ii) infectiveness in identifying gross polluting vehicles (Calvert et al., 1993; Bishop et al., 1997). Lack of proper enforcement, and corruption, prevents the realization of the full potential of any I/M program. Moreover, the lack of capacity, such as the lack of training of personnel, and poor quality test equipment, can hinder the success of the program. India is a classic example of how the lack of a well-conceived program defeats the overall objectives of the program [USAID, 2004]3. In Nepal, between 16–32 percent of vehicles failed the emissions test from 2000–2002 (Faiz et al., 2006). In Chongqing, China, only 10 percent of vehicles brought in by drivers failed the emissions test, but 40 percent of vehicles flagged down by roadside inspectors did not pass the emission test (USAID, 2004).

---

3 Indian I/M programs are plagued by poor quality personnel and test equipment, low compliance rates, and corruption. I/M tests are not taken by more than 15 percent of drivers and those who take it pass without truly controlling their emissions (USAID, 2004).
7. Other Laws and Regulations

Although policies such as fuel economy standards, emission standards, fuel quality standards, and I/M programs are most frequently utilized, they are by no means the only regulatory instruments introduced to discourage travel demand and reduce emissions from the transport sector. Several other regulatory measures have been experimented with, to varying degrees of success. For example, access bans, or partial and total vehicle bans, have been widely used in European countries such as Italy, Greece, The Netherlands, Spain, and Germany (Goddard, 1997). Italy has adopted a policy that bans private cars from entering the city centers. Italy aims to protect its historical city centers by not allowing non-residents to drive into the city center. In Swiss cities such as Bern and Zurich, the restrictive measures taken by the government (e.g., limited parking, road capacity reduction and diversion of through traffic) has made driving so difficult that many Swiss prefer using public transport (Bonnel, 1995).

The “No-Driving Day” (NDD) (or Hoy No Circula) policy introduced in Mexico City in 1989 is one of the much-discussed regulatory measures to control traffic congestion and vehicular emissions. It would not only help reduce environmental externalities through travel management but also reduce traffic congestion (Molina and Molina, 2004). The program mandates not driving one day during the week (except the weekends) and two days during serious pollution episodes. During the weekends, odd and even license plate numbers are used, which forced one-half of the fleet to be parked. By removing 20 percent of the vehicles from the streets in its first few months of operation, it did contribute towards the betterment of ambient air quality (Goddard, 1997). The gains made, however, were temporary. The program did not yield the desired level of success for several reasons. First, the city lacked sufficient public transport systems to meet the travel demand resulting from the ban on personal vehicles. Second, the driving public intelligently subverted the existing regulation. For example, many drivers adjusted to the restriction by purchasing additional autos in order to have at least one vehicle available on any given day. Many of the second vehicles were older and released more emissions. Some studies (e.g., Eskeland, 1994; Eskeland and Feyzioglu, 1995; Goddard, 1997) even argued that the program actually may have led to an increase in the number of vehicles and total emissions from road transport.

In addition to Mexico City, the traffic restriction (restricción vehicular) policy has been implemented in three different Latin American cities: Santiago (Chile), São Paulo (Brazil), and Bogotá (Colombia), with varying degree of success. The traffic restriction policy in Santiago implemented to reduce congestion and air pollution has limited the circulation of 20% of buses, taxis, and cars. In order to combat free rider problem, the schedule for the restriction is changed every few months. In São Paulo, the effects of the traffic ban have been undermined by growing car ownership. In order to meet air quality targets, Mexico City authorities are planning modifications to the existing scheme to ensure stricter enforcement with fewer exemptions (Mahendra, 2007).

Many Latin American countries have imposed complete or partial bans on used vehicles imports. Despite a huge market for used vehicles, countries such as Argentina, Brazil, Chile, Colombia, Ecuador, Mexico, Paraguay, Uruguay, and Venezuela have completely banned used vehicle imports (Pelletiere and Reinert, 2002).

The Supreme Court of India has played a proactive role in controlling vehicular pollution in New Delhi. Its directives include: (i) the phasing out of commercial/transport vehicles older than 15 years; (ii) the replacement of all pre-1990
autos and taxis with new vehicles using clean fuel; and, (iii) the conversion of the entire city bus fleet, both public and private, to use compressed natural gas (CNG) (DOT, 2009). The Supreme Court order for the conversion of the entire diesel-powered bus fleet in Delhi and its successful implementation clearly shows that reluctance on the part of the government in developing countries in maintaining air quality can be overcome through the judicial system.

In many large U.S. cities, regulation such as high-occupancy vehicle (HOV) lanes and high-occupancy toll (HOT) lanes have been introduced. These regulations help reduce emissions in two ways: (i) encouraging an increased vehicle occupancy and (ii) encouraging the use of clean vehicles (hybrids) and vehicles with higher fuel efficiency (motor cycles) as such vehicles are allowed in HOV and HOT lanes. Until recently, ten US States have considered allowing single occupant hybrid vehicles (SOHV) into HOV lanes (Chu et al., 2007). Although well intentioned, allowing hybrid vehicles in HOV lanes have started to produce negative externality in the form of increased congestion in HOV lanes. For example in Virginia, USA, where motorcycles and hybrid vehicles are allowed to take advantage of HOV facilities statewide, traffic congestion experienced by the commuters. In a survey conducted in 2002, vanpoolers cited Congestion in HOV lanes as their second greatest concern has been increasing (Poole and Balaker, 2005). Increased congestion also means increase in pollution.

8. Which regulatory instruments and where?

The literature on the design of regulatory policies to reduce transport sector externalities mainly focus on two central questions: (i) the desired level of protection of public health and environmental quality that a country or region is aiming to achieve and (ii) the cost and institutional capacity to implement the policies. Based on intent of the program(s), easing congestion or controlling pollution, the appropriateness of the regulatory instrument(s) under consideration may vary considerably. Factors that influence the effectiveness of the instrument should be used to gauge the appropriateness of the regulatory measure (Ghose, 2002; Satyanarayana, 2007). The selection of an instrument does not guarantee its effectiveness. The success of the selected instrument relies on factors such as: (i) the overall costs of emission control; (ii) the comprehensiveness of the law/regulations with regard to the level of development of the society; (iii) the ability of the industry in question to bear the control cost burden; and, (iv) the punitive measures in place and the chances of detection of violation (Priyadarshini and Gupta, 2003).

The choice of control options is based on the country’s priorities, the characteristics of the air pollution problem and the resources of the regulating agency (Cohen and Kamieniecki, 1991; Faiz and de Larderel, 1993). Take countries or cities facing severe local air pollution problems, for example. Most developing countries normally introduce emissions standards, whereas developed countries, which are equally concerned with both global and local air pollution, adopt a myriad of regulatory measures, such as fuel economy and fuel standards, in addition to emission standards. Since most developing countries are particularly concerned about local air pollution, they tend to prioritize the introduction of emissions standards and fuel quality standards over fuel economy standards. Moreover, implementation of emissions and fuel quality standards is
technologically simple with greater certainty of the desired outcome. In addition, except for few developing countries such as India, China, and Brazil which have automobile manufacturing plants and where setting fuel economy standards is more desirable, others are net importers of automobile and therefore they do not have control over the fuel economy standards.\(^4\)

Regulatory standards vary considerably from one country to another depending upon the level of motorization, dependency on private vehicles, and environmental consciousness. Fuel economy standards across European countries and between the United States and the EU vary significantly. Most developing countries are found to be reluctant to introduce stringent regulatory standards because of their limited resources to enforce the stringent standards (Cohen and Kamienicki, 1991; Priyadarshini and Gupta, 2003; Delfin, 2004).

Note here that regulatory standards are not mutually exclusive to each other in that introduction of an instrument does not obviate the need for others. For example, emission standards are necessary to control local air pollutants such as CO, HC, NO\(_x\), and fine particulate matter. Control devices reducing these emissions do not necessarily reduce fuel consumption and CO\(_2\) emissions and, hence, emission standards do not replace fuel economy standards. Similarly, emission standards may not replace fuel quality standards.

As the level of air pollution varies from one city to another, depending upon the level of motorization, compactness of the city, and maintenance level of the existing vehicle stock, most developing countries are struggling to make the selection of appropriate regulatory instruments that can effectively reduce emissions from the transport sector. One of the major questions, whose answer seems elusive for most, is what is the starting point in terms of framing effective policy instruments in reducing transport sector emissions? There seems to be no clear-cut answer to this question. There are, however, several worthy suggestions (Gwalliam, 2004; Mage and Walsh, 1992; ADB, 2003; Blumberg et al., 2003).

Understanding the factors affecting the total inventory of motor vehicle emissions is necessary to design effective programs. The ADB (2003) suggested that countries with a serious air pollution problem strongly consider leapfrogging to the most stringent standards possible, such as the Euro 2, Euro 3 or Euro 4, after making sure that the appropriate fuel is available. Blumberg et al. (2003) argued that jumping to near-zero sulfur diesel in a single step is more cost-effective and advantageous. The suggestions, although genuine, may not be always feasible due to the lack of resources, trained labor, and the required infrastructure. For example, one of the major difficulties associated with vehicle emission control programs is that it imposes significant economic and social costs (Gwalliam, 2004) and the actual beneficiaries are hard to identify (Faiz et al., 1999).

Motor vehicle pollution control programs should be based on a realistic assessment of costs and benefits and must be compared with the technical and administrative feasibility of proposed countermeasures. In order to make services affordable to the poor, transport policy must be designed to be both environmentally sensitive and consistent with public and private affordability.

\(^4\) However, they can impose import restrictions for low fuel economy vehicles.
9. Conclusions

This study reviews the main regulatory policy instruments to control transport sector externalities. The instruments considered include fuel economy standards, emission standards, fuel quality standards and other laws and regulations. We also highlight factors affecting the selection of regulatory instruments.

Fuel economy standards have generally been introduced in developed countries, which are not only concerned about local air pollutions but also other factors such as traffic congestion, climate change, and energy security. In the United States, fuel economy standards were first introduced in the early 1970s in an effort to lessen the impacts of the first oil crisis. Currently, the policy also serves to reduce GHG emissions. The fuel economy standard in the U.S. has not improved, however, since the 1985 level of 27.5 MPG, although the 2007 Energy Bill mandates an improvement to 35 MPG by 2020. In contrast to United States, the EU has defined fuel economy in terms of GHG emissions due to the increasing contribution of urban transportation to global GHG emissions. Implementation of the EU fuel economy standards will result in the reduction of vehicular CO\textsubscript{2} emission from 186 g/km in 1995 to 140 g/km in 2008 and further to 120 g/km by 2012.

Although the fuel economy standard is one of the key regulatory instruments employed in industrialized countries to reduce transport sector externalities, its success has been contested. Some existing literature argue that equivalent fiscal policy instruments, such as fuel or emission taxes, could have produced better results than fuel economy standards while reducing the same amount of fuel consumption and emissions. While the fuel economy standards help reduce fuel consumption and associated emissions, particularly CO\textsubscript{2} emissions, they do not necessarily reduce local and regional air pollutants, such as CO, VOC, NOx, and SPM to the level necessary to meet local air quality standards in many cities around the world.

Emission standards have been introduced in both industrialized and developing countries to control local air pollution. In response to the increase in local pollution level, vehicle emission standards have consistently been tightened over the years. Starting in 2004, tailpipe emissions standard for NOx has been set at 0.07 grams per mile in the U.S. (compared to 3.1 grams per mile in 1975). In the EU, there have been quick revisions in the emission standards towards advanced standards. The Euro 1 standards introduced in the early 1990s were modified to Euro 2 in 1996, to Euro 3 in 2000 and finally to Euro 4 in 2005. Following the footsteps of the industrialized countries, developing countries too have made commendable progress in terms of adopting emission standards. Several countries in Latin America and Asia have adopted either Euro or U.S. emission standards to control their local air pollution.

In order to control some pollutants, such as lead and oxides of sulfur, the element causing these pollutants needs to be limited through fuel quality standards. Most countries around the world have phased out leaded gasoline and controlled lead content in unleaded gasoline. Similarly, many countries, both industrialized and developing, have introduced fuel quality standards to limit sulfur content, thereby reducing oxides of sulfur and particulate matter. Moreover, several countries have introduced mandates for blending ethanol and biodiesel into, respectively, gasoline and diesel. This would certainly help reduce CO\textsubscript{2} and some local air pollutants.

Setting vehicular standards does not necessarily control emissions unless an effective enforcement mechanism is in place, however. Inspection and maintenance (I/M)
programs are the most common initiatives countries have undertaken to enforce the standards. The programs mandate regular inspection of vehicles and retirement of those not meeting the standards. Besides standards, there also exist some regulatory measures, such as import ban on polluting vehicles in many Latin American countries, partial and complete driving restrictions in some European cities, the no driving day program in Mexico City and the mandatory conversion of public bus in New Delhi from diesel to compressed natural gas (CNG).

Fuel economy standards, emission standards, fuel quality standards and I/M programs are not mutually exclusive and they are introduced for different purposes. Different countries could give priority to different measures depending upon their needs and institutional capacity to enforce the standards. Since most developing countries are particularly concerned about local air pollution, they are found to prioritize the introduction of emissions standards and fuel quality standards over fuel economy standards.

Acknowledgements

We sincerely thank Asif Faiz, Zachary Moore, Mike Toman, Ashish Shrestha and an anonymous referee for their valuable comments and suggestions. We acknowledge the Knowledge for Change Program (KCP) Trust Fund for the financial support. The views expressed in this paper are those of the authors and do not necessarily represent the World Bank and its affiliated organizations.
References


Appendix

Acronyms

CO  Carbon Monoxide  
EPCA  Energy Policy and Conservation Act  
EPCA  European Automobile Manufacturers Association  
EC  European Commission  
EU  European Union  
GHG  Greenhouse gas  
GVWR  Gross Vehicle Weight Rating  
HC  Hydrocarbon  
JAMA  Japanese Manufacturers  
KAMA  Korean Manufacturers  
MPG  Miles Per Gallon  
NAFC  National Average Fuel Consumption  
NMOC  Non-metallic Organic Compounds  
NOx  Oxides of Nitrogen  
SOx  Oxides of Sulfur  
SPM  Suspended particulate matters  
CAFE  The Corporate Average Fuel Economy  
FCAI  The Federal Chamber of Automotive Industries  
VMT  Vehicle Miles Traveled  
VOCs  Volatile Organic Compounds